KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Passenger robust timetables for dense railway networks

Sofie Burggraeve

Supervisor:
Prof. dr. ir. P. Vansteenwegen

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Mechanical Engineering

September 2017

Sofie BURGGRAEVE

Examination committee:
Prof. dr. A. Bultheel, chair
Prof. dr. ir. P. Vansteenwegen, supervisor
Prof. dr. F. Spieksma
Prof. dr. ir. D. Cattrysse
Prof. dr. R.M. Lusby
(DTU Management Engineering, Denmark)
Prof. dr. M. Schmidt
(Erasmus University Rotterdam, The Netherlands)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Mechanical Engineering

Cover image: courtesy of Marcel Derweduwen, treinreiziger.nl

September 2017

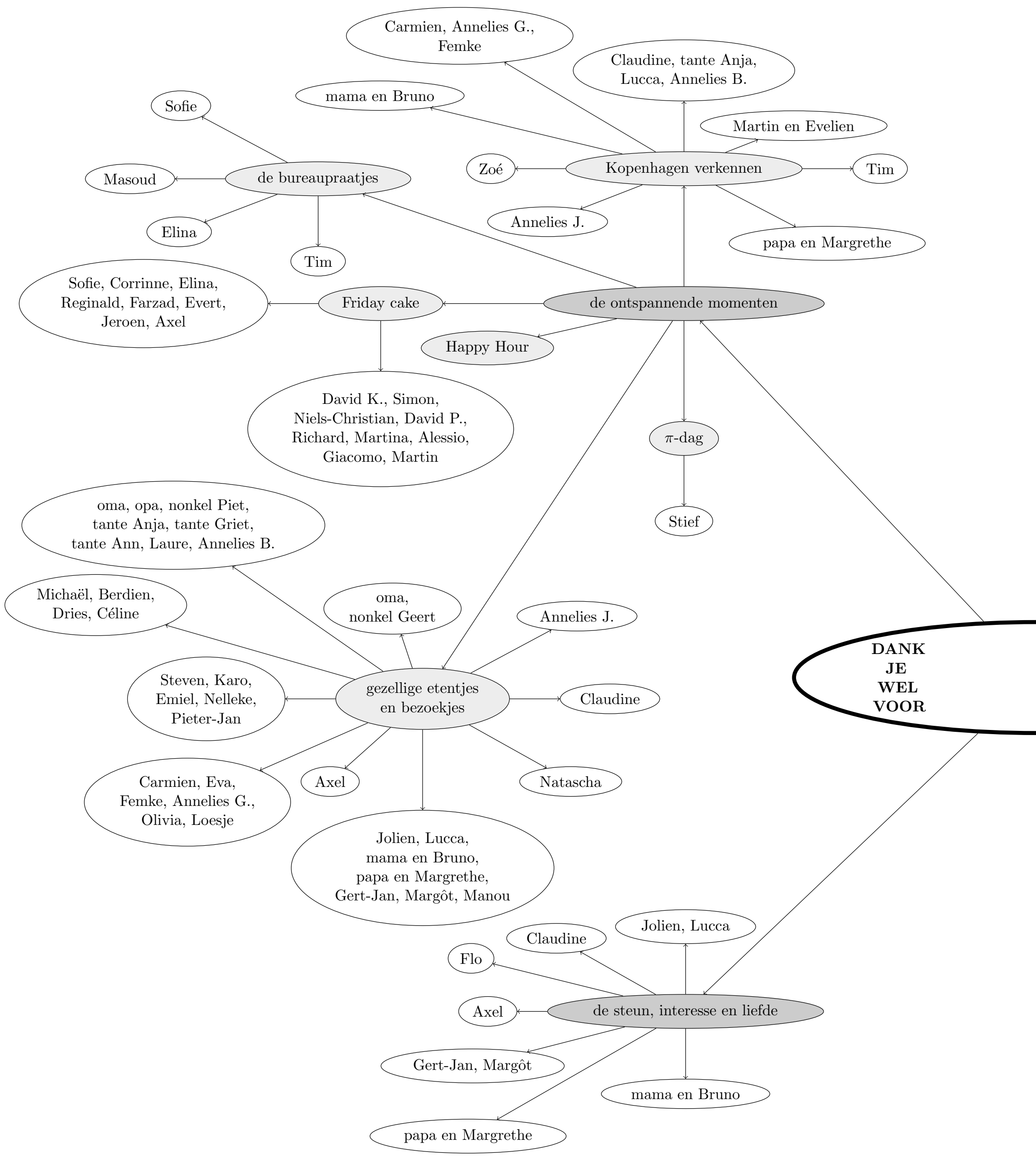
Uitgegeven in eigen beheer, Sofie Burggraeve, Celestijnenlaan 300A box 2422, B-3001 Leuven (Belgium)

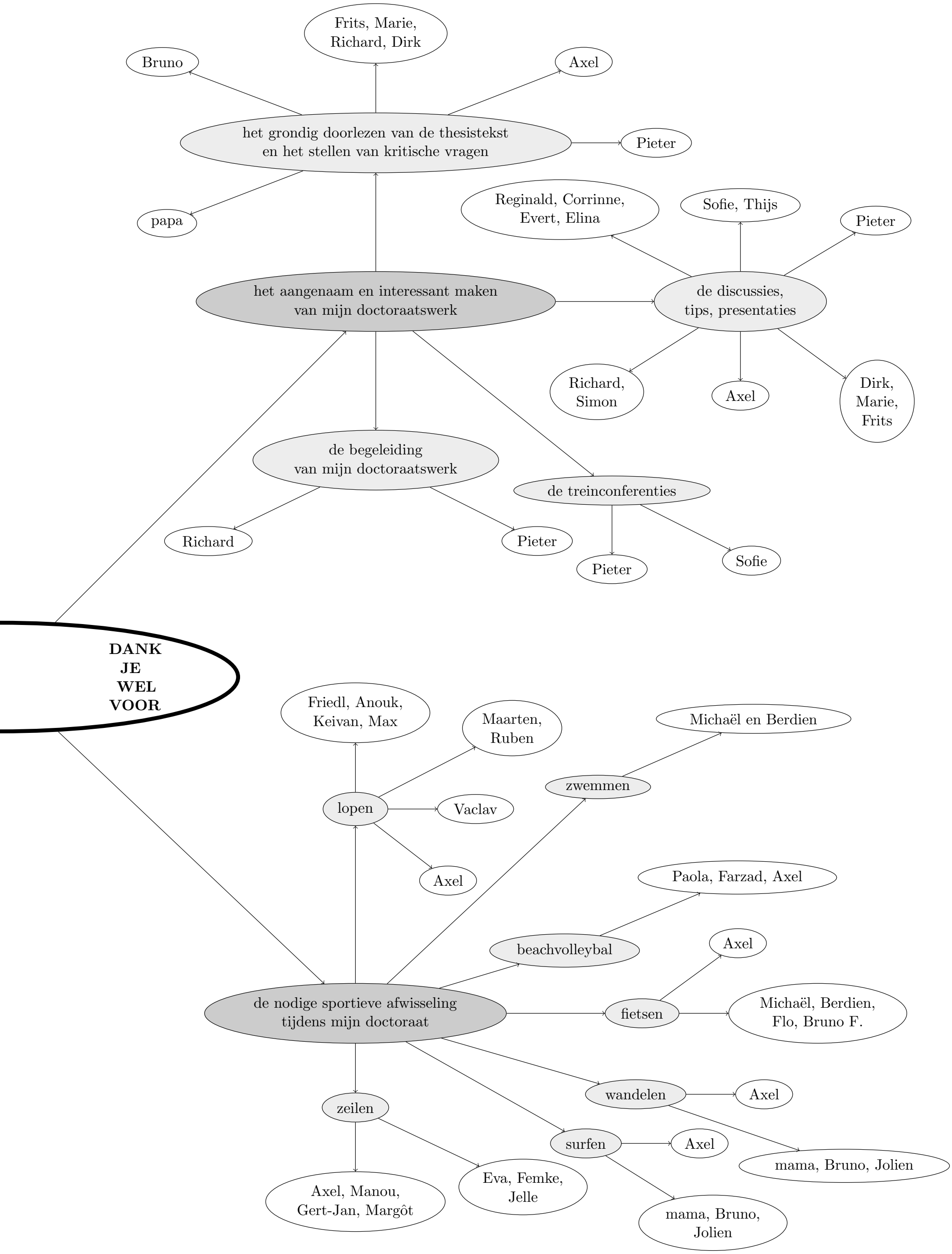
Preface

Dit boekje is het resultaat van mijn eerste vier jaar als afgestudeerde wiskundige. Het beschrijft mijn doctoraatsonderzoek dat ik gedaan heb in de me ondertussen meer dan bekende ingenieursomgeving. Deze doctoraatsjaren waren soms intensief, soms ontnuchterend, maar ik had heel veel mensen rondom mij die me elk op een andere manier gesteund en geholpen hebben: bij het opdoen van nieuwe kennis, bij het uitwisselen van ideeën, in mij te begeleiden, bij de sportieve en de ontspannende activiteiten als afwisseling tussen het nadenk- en programmeerwerk of in de onvoorwaardelijke steun die ze me hebben geboden.

Een uitgebreid dankwoord schrijven is niet voor me weggelegd, maar de mindmap op de volgende twee pagina's moet dit compenseren. Iemand die ik nog in het bijzonder wil bedanken is Pieter omdat hij me steeds de nodige ruimte heeft gegeven om mijn eigen weg te kiezen in mijn doctoraat en tegelijkertijd de nodige eisen stelde. Hij heeft ervoor gezorgd dat er tactisch de juiste beslissingen werden gemaakt om dit doctoraat tot een goed einde te brengen.

Axel, ik kende jou niet toen ik aan mijn doctoraat begon, maar onze ontmoeting maakt de afgelopen vier jaar al meer dan de moeite waard. Ik heb er van genoten om dagelijks met jou naar het werk te kunnen fietsen en om je steeds dichtbij te hebben. Bovendien valt jouw bijdrage aan het behalen van mijn doctorstitel niet te onderschatten. Jij slaagt erin om mijn focus te verleggen van de uitzonderingsgevallen waar het misloopt naar de sterke punten in mijn verhaal. Dank je wel mijn lieve schat!


Carmien, Annelies G., Femke
Claudine, tante Anja, Lucca, Annelies B.
mama en Bruno
Sofie
Martin en Evelien
Masoud
de bureaupraatjes
Zoé
Kopenhagen verkennen
Tim
Annelies J.
Elina
papa en Margrethe
Tim
Sofie, Corrinne, Elina, Reginald, Farzad, Evert, Jeroen, Axel
Friday cake
de ontspannende momenten
Happy Hour
David K., Simon, Niels-Christian, David P., Richard, Martina, Alessio, Giacomo, Martin
π-dag
Stief
oma, opa, nonkel Piet, tante Anja, tante Griet, tante Ann, Laure, Annelies B.
Michaël, Berdien, Dries, Céline
oma, nonkel Geert
Annelies J.
DANK JE WEL VOOR
Steven, Karo, Emiel, Nelleke, Pieter-Jan
gezellige etentjes en bezoekjes
Claudine
Carmien, Eva, Femke, Annelies G., Olivia, Loesje
Axel
Natascha
Jolien, Lucca, mama en Bruno, papa en Margrethe, Gert-Jan, Margôt, Manou
Claudine
Jolien, Lucca
Flo
Axel
de steun, interesse en liefde
Gert-Jan, Margôt
mama en Bruno
papa en Margrethe



DANK JE WEL VOOR
het aangenaam en interessant maken van mijn doctoraatswerk
het grondig doorlezen van de thesistekst en het stellen van kritische vragen
Bruno
Frits, Marie, Richard, Dirk
Axel
Pieter
papa
de discussies, tips, presentaties
Reginald, Corrinne, Evert, Elina
Sofie, Thijs
Pieter
Richard, Simon
Axel
Dirk, Marie, Frits
de begeleiding van mijn doctoraatswerk
Richard
Pieter
de treinconferenties
Pieter
Sofie
de nodige sportieve afwisseling tijdens mijn doctoraat
lopen
Friedl, Anouk, Keivan, Max
Maarten, Ruben
Vaclav
Axel
zwemmen
Michaël en Berdien
beachvolleybal
Paola, Farzad, Axel
fietsen
Axel
Michaël, Berdien, Flo, Bruno F.
wandelen
Axel
mama, Bruno, Jolien
surfen
Axel
mama, Bruno, Jolien
zeilen
Axel, Manou, Gert-Jan, Margôt
Eva, Femke, Jelle

Abstract

Delays and unreliable travel times are daily practice and unavoidable in public transport. Furthermore, manual decisions are still omnipresent in planning processes and real time dispatching. Nevertheless, the planning of public transport affects the performance and the popularity of these transport modes and the sustainability of the transport system as a whole. Therefore, investigating how public transport can benefit from decision support to improve the planning seems a promising direction for future research.

This dissertation focuses on automatically developing timetables, routing plans and line plans for railway bottlenecks, such as the Brussels railway area (Belgium) and the Copenhagen S-tog network (Denmark). The connecting thread throughout this dissertation is providing a better service to passengers. Therefore passenger robustness is optimized, or in other words, the total travel time in practice of all passengers, in case of frequently occurring small delays, is minimized. Delay propagation from one train to another, which leads to unreliability and a lengthening of the passenger travel times, negatively affects this passenger robustness. Spreading the trains in time and space and including appropriate supplements in the timetable are solutions against delay propagation.

This dissertation presents several timetabling methods that incorporate these solutions against delay propagation in an intelligent way. Furthermore, these timetabling methods are either integrated with a routing model or with a line planning model. Both a timetabling method to construct a passenger robust timetable from scratch and a method to improve the passenger robustness of an existing schedule are provided.

All the methods are validated on real world case studies under varying circumstances. The computational results show that the developed methods are worthwhile to implement in practice. First because these methods allow to use the limited infrastructure in railway bottlenecks more efficiently, but even more so since simulation shows that these methods provide a much better service to the passengers than current practice. This state-of-the-art research is ready to be implemented in practice by railway companies, if they want to improve the service they offer to the passengers.

The best timetable and routing plan for Brussels, developed from scratch with the methods presented in this dissertation, improve the passenger robustness up to 17% compared to the Belgian railway infrastructure manager Infrabel and up to 8% compared to the best timetable and routing plan from the literature. For the DSB S-tog network of Copenhagen, the integrated method for line planning and timetabling constructs line plans and timetables from scratch for which the weighted sum of operator and passenger cost is close to the optimal weighted sum of these costs. Moreover, the passenger robustness highly improves compared to the initial line plans and timetables in eight out of the ten studied cases.

Beknopte samenvatting

Vertragingen en onbetrouwbare reistijden zijn helaas dagelijkse kost voor gebruikers van het openbaar vervoer. Bovendien worden er nog steeds veel manuele beslissingen genomen bij het plannen en het real-time bijsturen van bussen, treinen, trams, enz. Deze twee factoren zijn bepalend voor de performantie en dus de populariteit van het openbaar vervoer en beïnvloeden ook de duurzaamheid van het hele transportsysteem. Er is dus een grote nood om te onderzoeken hoe het openbaar vervoer beter gepland kan worden met behulp van een beslissingsondersteunend systeem.

Deze doctoraatsthesis spitst zich toe op het automatisch opstellen van dienstregelingen, routeplannen en lijnplannen voor spoorwegbottlenecks. Het stationsgebied in Brussel (België) en het S-tog netwerk in Kopenhagen (Denemarken) worden specifiek onder de loep genomen. De rode draad doorheen deze thesis is het streven naar een verbeterde service voor de passagiers. Om dat te bereiken wordt de passagiersrobuustheid geoptimaliseerd, wat betekent dat de totale reistijd van de passagiers in de praktijk geminimaliseerd wordt, rekening houdend met de kleine vertragingen die zich dagelijks voordoen. Treinen die vertragingen aan elkaar doorgeven zorgen voor onbetrouwbare en langere reistijden voor de passagiers. Dit heeft een negatief effect op de passagiersrobuustheid. Het spreiden van de treinen in de voorziene tijd en op de beschikbare infrastructuur en het invoegen van doeltreffende supplementen in de dienstregeling kunnen het doorgeven van vertragingen in belangrijke mate voorkomen.

Deze thesis presenteert verschillende methodes om dienstregelingen op te stellen die op een slimme manier gebruik maken van het spreiden van de treinen en het invoegen van supplementen om het doorgeven van vertragingen tegen te gaan. Deze methodes worden bovendien geïntegreerd met het opstellen van een routeplan of een lijnplan. De methodes kunnen zowel gebruikt worden om een passagiersrobuuste dienstregeling vanaf nul op te bouwen of om de passagiersrobuustheid van een bestaande dienstregeling te verbeteren.

Deze methodes zijn allemaal gevalideerd op een realistische gevalstudie onder variërende omstandigheden. De resultaten tonen aan dat het zinvol en de moeite waard is om de ontwikkelde methodes in de praktijk te implementeren. Eerst en vooral omdat deze methodes een efficiënter gebruik van de beperkte infrastructuur in een spoorwegbottleneck toelaten, maar vooral omdat deze methodes de service voor de passagiers verbeteren ten opzichte van de huidige praktijk. Dit state-of-the-art onderzoek kan zo in de praktijk bij spoorwegbedrijven geïmplementeerd worden, als zij de service die ze aan de passagiers leveren willen verbeteren.

In deze thesis worden een dienstregeling en routeplan voor de bottleneck in Brussel opgesteld die de passagiersrobuustheid verbeteren met meer dan 17% in vergelijking met de planning van de Belgische spoorweginfrastructuurbeheerder Infrabel en met meer dan 8% in vergelijking met de beste planning verkregen in voorafgaand wetenschappelijk onderzoek. Voor het DSB S-tog netwerk in Kopenhagen kan de geïntegreerde methode lijnplannen en dienstregelingen vanaf nul opstellen waarvan de gewogen operator- en passagierskost dichtbij het optimum ligt. In acht van de tien onderzochte gevallen verbeteren deze lijnplannen en dienstregelingen daarenboven ook de passagiersrobuustheid ten opzichte van de initiële lijnplannen en dienstregelingen.

Contents

List of Figures

List of Tables

1 Introduction

During the past four years, when I was at work on this dissertation, many people have expressed their interest in the research I was doing. After having told me about their many railway adventures, they invariably expressed the hope that I would be able to come up with railway time schedules that provide fast travel times and would be much more reliable than they are today. Some people are even dreaming about a delay free railway system... Their reasoning can actually easily be summarized as follows: they dislike delays, since (1) these cause them to arrive late, (2) these oblige them to take extra travel time into account when planning their train trip, as the schedule is perceived as unreliable and (3) these delays are the cause of missing a connecting train, which in turn gives cause to annoying waiting times in stations. Moreover they expressed their concern that the timetable accuracy they *perceive* seems to differ substantially from the accuracy measure that is regularly *put forward* by the management of the railway system. It flattered me that people put so much confidence in me and the research I was doing, and I always tried to make them happy by informing them about the progress of my research activities. However, I also had to disappoint them by stating clearly that it is plainly impossible to avoid all delays. More importantly however, these conversations gave me a continuous motivation, not only to put passenger service first on the list, but especially so in models that could actually be used in practice, also in the huge Belgian railway bottleneck in Brussels. Moreover, these conversations served as a justification to advocate models in which the main focus is on 'the passenger' and not so much on 'the train'. In order to keep the main focus on 'the passenger', I will not be able to resort to model solutions in which a substantial supplement time is scheduled somewhere near the end of a train line, in order to make the arrival time in the terminal station more reliable. This can however be observed in current practice and can be explained since today the punctuality is only measured just before Brussels and/or in the terminal station of the train. It is easy to see that this kind of accuracy will not guarantee optimal timetables for most passengers traveling between intermediate stations. My goal is therefore to develop ‘passenger robust’ time schedules and this will be the preferred criterion throughout this entire dissertation. In the end, a better performing railway system, characterized by a better passenger service, a higher accuracy, shorter trips and more reliable connections, could actually also convince commuters using other means of transport to use the train on a regular basis. This would increase the sustainability of the entire transport network in Belgium and reduce ever increasing congestion issues.

In order to improve the planning of train routes and the corresponding time schedules and to make these more passenger-friendly, appropriate mathematical methods are required. Even though this research area is very closely connected to real life issues, many useful models, results and proceedings developed in the context of railway research are unfortunately not yet implemented in the best current practice of railway planners. Railway planning is complex; given that many players interact whilst having different preferences, e.g. the commuters, the railway infrastructure manager, the railway operator, the crews, maintenance teams, etc. The planners have to make decisions over different time horizons, e.g. long term network adaptations, timetabling, real time dispatching, etc. Railway bottlenecks, which can be characterized by dense train traffic and a complex infrastructure lay-out, make the planning even more difficult. In such a bottleneck, a small disturbance caused by one train may easily affect many other trains. Furthermore, a lot of precise and correct data is needed to model the railway system in detail, such as the infrastructure lay-out, the passenger demand, the dispatching strategies, etc. This dissertation will present a lot of different mathematical methods that will be able to improve the planning at railway bottlenecks, whilst keeping the focus on the passenger in order to result in an improved quality of the public transport service.

1.1 Research goals

While mathematical concepts are at the core of railway planning, they are not always exploited to automate railway planning in practice. In many countries mathematical tools are used to facilitate the planning process, but often human experience is still essential. In recent years, robustness has also made its appearance in the planning process in practice. Actually, there are different interpretations and definitions of robustness. The one that is preferred and premised in this dissertation is *passenger robustness* (Dewilde et al., 2011). It differentiates itself from other definitions, since it aims for both reliable and short travel times in practice for passengers and it strives for reliable travel times for trains. Since delays cause longer and unreliable travel times, delays are inherently avoided by constructing a passenger robust timetable.

Railway research on timetabling is extensive and a lot of challenges are already covered. The aim of the existing research typically differs along the following axes: normative macroscopic versus microscopic, traditional optimization criterion (e.g. conflict-freeness, minimum planned travel times, etc.) versus (passenger) robustness, small network versus large network, sparse or moderately dense network versus very dense network, improving an existing timetable versus starting from scratch. The research in this dissertation concentrates on the latter option in each combination.

The objective of this dissertation is **to create mathematical models to construct a passenger robust timetable for dense railway networks**. The focus on passenger robustness in dense railway networks is as yet barely explored, since it is already a challenge to construct a feasible conflict-free planning for such railway networks. But in fact, passenger robustness is indispensable in dense railway networks in order to assure the assigned time and the reliability of the planned travel times. An additional objective in this dissertation is to create a passenger robust schedule for several planning steps. Therefore the integration of different planning steps is investigated in this dissertation, namely the interaction between the construction of the routing plan and the timetable and the interaction between the construction of the line plan and the timetable.

1.2 Scope of the thesis

The railway research in this dissertation is restricted to dense and complex railway networks, since these are challenging and have hardly been studied in literature. A first example, explicitly studied in this dissertation, is the timetabling and routing of 85 trains through the Brussels bottleneck with its 28 970 248 alternative routes and a restricted number of tracks at the center of this bottleneck. A second example, which is also explicitly studied in this dissertation, is the timetabling and line planning of sixty trains through a corridor with only one platform in each direction. Furthermore, the trains have different terminal stations and no shunt capacity is available in these terminal stations. It goes without saying that both a complex infrastructure lay-out and a dense occupancy of trains make the planning complicated, but crucial for the performance of the whole network. Dense railway areas are actually most vulnerable to delays. Therefore it is useful to focus on the planning of a bottleneck before planning the rest of the network (Goldratt and Cox, 1986). Thereafter the planning can generally be extended and made feasible for the whole network outside this bottleneck without many changes (to this planning), since typically many fewer constraints are present outside the bottleneck.

Furthermore, the research in this dissertation is focused on schedules for passenger trains during peak hour. The motivation here is that freight trains are mostly avoided in bottlenecks during peak hour. This will also be the case in both dense networks considered in this dissertation, namely the railway bottleneck in Brussels (Belgium) and the high frequency S-tog network in Copenhagen (Denmark).

Railway planning consists of several subproblems. These subproblems can be classified according to their time horizon. In the long term, i.e. the strategic level, decisions are made about the railway infrastructure, for example new network parts can be constructed or useless infrastructure can be removed. The same holds for line planning, where decisions are made on the origins, destinations, frequencies and stops of the trains, and which is typically made for several years. In the medium term, i.e. at the tactical level, decisions are made about the timetable, the rolling stock, the routing plan and the crew. In the short term, i.e. at the operational level, real time dispatching decisions have to be made. An overview and the interactions of the subproblems present in railway planning can be found in Figure 1.1. This overview is discussed in more detail in Section 2.1.3.

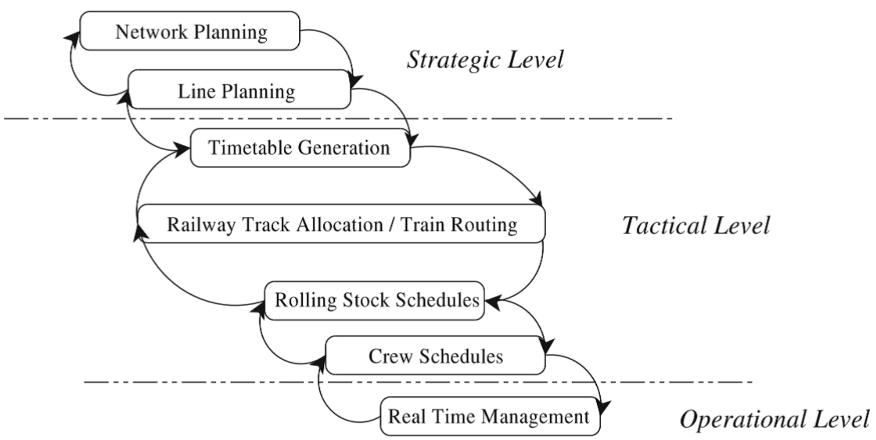


Figure 1.1: Railway planning steps (Lusby et al., 2011).

This dissertation covers railway timetabling at the tactical level, the interaction between railway timetabling and routing and the interaction between railway timetabling and line planning. The routing problem and the line planning problem are handled separately from the timetabling problem but only in function of their interaction with timetabling. This differentiates the solutions to both the routing and the line planning problem proposed in this thesis compared to other solution approaches for these problems. The methodology to construct the routing plan spreads the trains over the available infrastructure. Consequently the interactions between trains are minimized, which is advantageous for timetabling. The methodology to construct the line plan provides flexibility in the timetabling process in order to avoid knock-on delays between trains. Network planning, rolling stock scheduling, crew scheduling and real time dispatching are not investigated throughout this dissertation.

Large disruptions in dense networks will seriously disturb any timetable. In contrast, the impact of frequently occurring small delays can be avoided by constructing a passenger robust timetable. That is why the focus in this dissertation is on these frequently occurring small delays while constructing the timetable and also while testing the performance of the timetable in a simulation. However, the proposed approach may also be useful in case of larger disruptions. The delay data that is used in this dissertation originates from the Belgian railway infrastructure manager Infrabel and has been processed to functional data by dr. Peter Sels and dr. Thijs Dewilde.

1.3 Contributions of the thesis

The first contribution is the improvement of the methodology presented in Dewilde et al. (2013, 2014a). This latter methodology focuses on a railway bottleneck. It starts from an existing timetable, routing plan and platform assignment. By iteratively making small changes to each part of the schedule, the passenger robustness of the schedule is improved, while assuring microscopically conflict-freeness. The contribution in this dissertation concerns explicitly taking into account passenger numbers and recurring delays in the methodology of Dewilde et al. (2013, 2014a). The free time between two trains is weighted based on the passengers on the later train and/or on the expected delays of both trains. This leads to a spreading of the trains for which potential conflicts that affect many passengers are avoided. This is beneficial for the total travel time in practice of all passengers. Using this approach, the passenger robustness of a reference timetable and routing plan for Brussels, Belgium's main railway bottleneck, produced by the Belgian railway infrastructure manager Infrabel, can be improved by 11%. This also constitutes an improvement of almost 4% compared to the best performing routing plan and timetable from the literature for Brussels (Dewilde et al., 2013, 2014a). The framework in which this contribution is achieved is described in Chapter 3.

The second contribution is the development of a mathematical model for the construction of a microscopic conflict-free passenger robust timetable and routing plan. First, every train is assigned to a route such that the maximum node usage is minimized and that the number of times that each node is used, is quadratically penalized. Thereafter the blocking times that indicate when the nodes on the trains' routes are reserved and released, are assigned. The difference with other approaches is that we focus on the occupation of the railway infrastructure before constructing the timetable. In a limited amount of time, the mathematical model constructs a conflict-free routing plan and timetable from scratch and achieves a passenger robustness which is also up to 11% better compared to the passenger robustness of the reference timetable and routing plan from Infrabel. This is also an improvement of 2% compared to the best performing routing plan and timetable from literature for Brussels (Dewilde et al., 2013, 2014a). This achievement is explained in Chapter 4.

The third contribution is the development of an iterative methodology in which supplements and buffer times are cleverly weighed off against each other in order to create a microscopic passenger robust timetable. Each iteration is based on four pillars, which are subsequently executed. The first two pillars are comprised by the routing and the timetabling model which are already treated in the second contribution. The third pillar consists of a simulation, which is used to evaluate the passenger robustness of the routing plan and the timetable. In the fourth pillar, the simulation outcome is used to make a new supplement assignment for the next iteration. The methodology is tested on the same case study for Brussels. The passenger robustness is up to 17% better than that of the reference timetable and routing plan from Infrabel. This is an improvement of 8% compared to the best performing routing plan and timetable from literature for Brussels (Dewilde et al., 2013, 2014a). The details of this achievement can be found in Chapter 4.

The fourth contribution is the development of a heuristic algorithm to build a railway line plan from scratch, that minimizes passenger travel time and operator cost and for which a feasible and passenger robust timetable exists. A line planning module and a timetabling module work iteratively and interactively. Flexibility created in the line planning module is used during timetabling to improve the robustness of the railway system. The algorithm is validated on the DSB S-tog network of Copenhagen. While the operator and passenger cost remain close to the costs for which the initial line plan had been optimized, the timetable corresponding to the finally developed robust line plan significantly improves the passenger robustness, in eight out of ten studied cases. The details of this achievement can be found in Chapter 5.

1.4 Practical relevance of the thesis

The achievements presented in this dissertation can be implemented in practice. The method proposed in the first contribution can be applied to any network, timetable and routing plan to improve the passenger robustness. It uses a timetable and routing plan as input and only makes small adjustments that improve the passenger robustness, while maintaining the conflict-freeness. Specific requirements of trains, passengers or infrastructure have to be checked while making these adjustments. The profit will be larger for dense and restricted networks for which accurate passenger numbers and delay data are known.

For the second contribution it holds that the most recent data on the line plan and the present network are needed to construct a passenger robust timetable from scratch with the timetabling model. Extra constraints to model specific requirements of trains, passengers or infrastructure can be added to the model without structurally changing the model. Unfortunately, more or stricter constraints, reduce the flexibility for including passenger robustness in the schedule. The routing model needs an overview of all routes between any two considered stations. To create this overview, efficient algorithms are present in the literature. This overview has to be made only once for each network. If the infrastructure changes locally, the overview has to be adapted for that region. To create a planning for the whole Belgian railway network, the routing and the timetabling model have to be accompanied by a methodology to expand the timetable to the rest of the network and to plan trains that do not cross Brussels. Since typically fewer constraints are present outside the bottleneck, this expansion is supposed to barely affect the schedule for the bottleneck. The routing model and the timetabling model start from the building blocks of a railway system, namely signals, platforms, switches and blocking times, which makes these models generally applicable to all kind of railway networks. However, the larger and more complex the network, the more computation time will be needed. The iterative methodology considered as the third contribution is an extended version of the former timetabling and routing model and is very promising. Targeted fine tuning will broaden its applicability and the generated profit of using this integrated approach. The approach is based on the same data as for the timetabling and routing model, but also needs passenger numbers and historical delay data.

Concerning the fourth contribution, the resulting timetable and line plan can be immediately implemented in practice on the DSB S-tog network of Copenhagen. The created line plan is near cost-optimal and fulfills all service constraints. The created timetable satisfies the turning constraints in the terminal stations and is macroscopic conflict-free. This means that all the macroscopic norms (e.g. headways and occupation times) are satisfied. If passengers would be open to line plans which contain similar lines but different stopping patterns, even better schedules could be created with the approach. Presently, relocation of signals and placing of extra signals is going on on the DSB S-tog network in order to shorten the sections. This has the consequence that the blocking times of sections will decrease. By using the (new) data on the signal locations and the timetabling model referred to in the second and third contribution, a microscopic timetable can be designed which is more accurate and assures conflict-freeness at the microscopic level. This may also lead to a more passenger robust schedule.

A lot of S-line, tram and metro networks have a similar structure as the DSB S-tog network of Copenhagen. With the necessary and correct data on connected stations, train travel times, transfer stations, passenger demand, safety and service requirements and turn restrictions, the integrated method, comprising the fourth contribution, can immediately be applied to construct a cost-optimal line plan and accompanying passenger robust timetable for these networks.

1.5 Outline

The remainder of this thesis is organized as follows. Chapter 2 defines all terminology used in this dissertation. Furthermore, it discusses the state of the art of railway timetabling, routing, line planning and integrated methods. It situates the dissertation in the literature by handling similarities to and differences with existing approaches. It also contains a separate section on how robustness could be implemented in the different planning stages.

Chapter 3 presents a methodology to include passenger robustness in an existing timetable. Passenger numbers and recurring delays are explicitly taken into account. The methodology is validated on the main railway bottleneck of Belgium, Brussels.

Chapter 4 presents an integrated approach for the construction of a microscopic passenger robust routing plan and timetable. The methodology looks for a globally optimal solution while spreading the trains in time and space. This methodology is also validated on the railway station area of Brussels and the results are compared to a reference timetable and routing plan from the Belgian railway infrastructure manager Infrabel and the best timetable and routing plan from the literature for this set-up (Dewilde, 2014b). This chapter also contains a part in which it extends the approach by including an extra degree of freedom. Besides buffer times, also supplements are optimized in this extended approach in order to create an even more passenger robust schedule. The extended approach works iteratively whereby new supplements used in the next iteration are based on the simulation output of the timetable and routing plan of the current iteration.

Chapter 5 analyzes the interaction between railway line planning and timetabling. It presents a heuristic algorithm to build a railway line plan from scratch that minimizes passenger travel time and operator cost and for which a feasible and robust timetable exists. This chapter also explains how limited shunt capacity and certain frequency combinations of lines that share a part of the network can lead to timetable-infeasibility of line plans. A validation of the methodology on the DSB S-tog network of Copenhagen is also contained in this chapter.

Chapters 3 to 5 are based on internationally peer-reviewed papers.

Chapter 6 concludes the thesis. It summarizes the obtained results and proposes research questions for future research.

2 State of the art

This chapter starts by defining all concepts present in this dissertation. Then railway robustness is discussed and how it is implemented in practice. Thereafter the state of the art of railway routing, timetabling, line planning and integrated methods are discussed, together with the similarities to, the differences with and the new contributions of this dissertation. [1]

2.1 Definitions

First the basic elements that build up a railway network are defined. Then, it is defined how trains drive through a railway network. The third part of this

[1]This chapter contains the literature review sections of the peer-reviewed journal papers:

Sofie Burggraeve, Simon H. Bull, Pieter Vansteenwegen, and Richard M. Lusby. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160, 2017

and:

Sofie Burggraeve and Pieter Vansteenwegen. Robust routing and timetabling in complex railway stations. *Transportation Research Part B: Methodology*, 101:228–244, 2017c.

and:

Sofie Burggraeve and Pieter Vansteenwegen. Optimization of supplements and buffer times in passenger robust timetabling. *Journal of Rail Transport Planning & Management*, In Press, 2017b. URL `https://doi.org/10.1016/j.jrtpm.2017.08.004`.

Many sentences are cited from theses papers, but in order to keep the text easily readable, cited sentences are not separately indicated.

section defines the railway planning problems considered in this dissertation.

2.1.1 How is a railway network built up?

A railway network can be looked at from a macroscopic and a microscopic point of view. The microscopic level is much more detailed and especially relevant to train drivers and dispatchers, while passengers usually only come in contact with the macroscopic level.

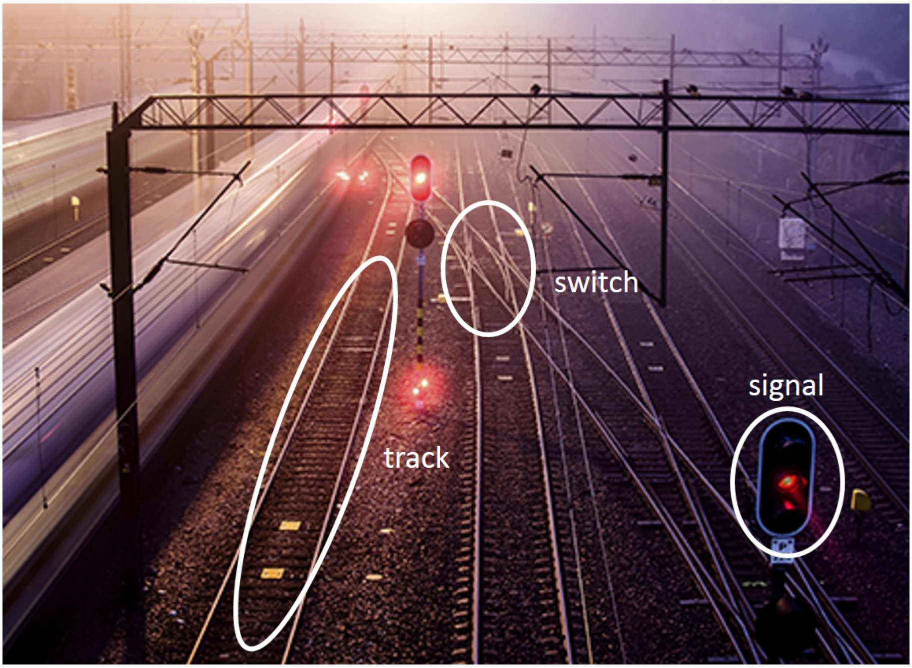


Figure 2.1: Tracks, switches, signals, platforms and border nodes are the building blocks of a railway network. The latter two are not visible on this figure (Source: Ricardo Rail).

At the microscopic level, a railway network is built up by tracks, switches, platforms, signals and border nodes. A real representation of tracks, signals and a switch can be seen in Figure 2.1[2]. *Border nodes* are the track ends or platforms at the border of the considered network. Most tracks, switches and platforms can be used in both directions, but sometimes they are permanently assigned to trains coming from only one direction. *Signals* give information to

[2]Source: https://magnet.me/a/company/ricardo-rail/about

trains coming from the direction from which the signal is visible. A railway network contains signals to safely guide the trains through the network, since at higher speeds the braking distance of a train is longer than the train driver's field of view. This implies that, at higher speeds, it is already too late to avoid a crash or accident at the instant that the train driver observes the obstacle. A *switch* enables a train either to be guided from one track to another track or to cross another track. It is either a junction to a parallel track or siding or between two crossing tracks. Every pair of consecutive similarly directed signals determines a *section*. The tracks, switches and platforms between these signals belong to this section.

A microscopic railway network can be represented by a graph where switches, platforms and network border nodes are nodes and the tracks between these nodes are the links. From now on we will refer to switches, platforms and border nodes as *nodes*. An example of a graph representing a microscopic railway network can be found in Figure 2.2.

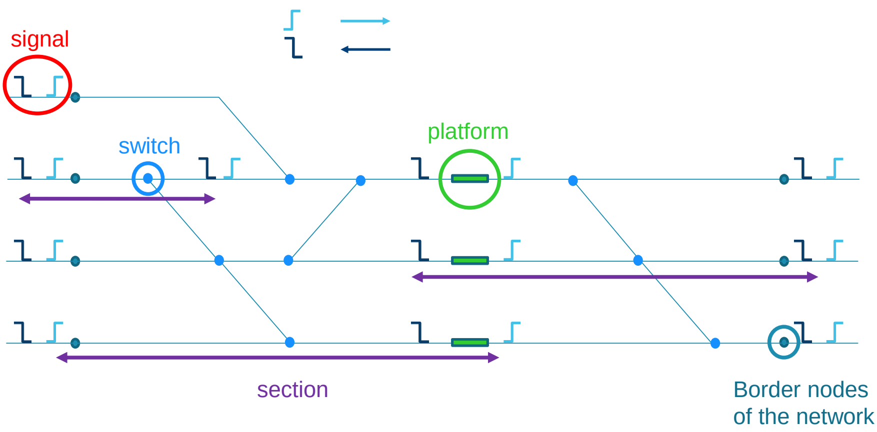


Figure 2.2: Representation of a microscopic railway network.

At the macroscopic level a railway network is built up by stations, corridors and grid zones. *Stations* are the parts of the network containing one or more parallel platforms. Here, passengers can enter and leave the train or transfer to another train. *Grid zones* are those parts of the network where trains can reach multiple parallel tracks or have access to several platforms due to an accumulation of switches. Grid zones are mostly located next to well-connected stations. In a grid zone, trains are guided to the track leading to their next stop. *Corridors* are the remaining parts of the network. They consist mainly of one or two parallel tracks in each direction, but can also consist of a single track which must be shared between trains in two directions. A corridor can also contain a low number of switches (compared to a grid zone) connecting parallel tracks.

A macroscopic railway network can be represented by a graph where stations are the vertices and an edge represents a corridor or grid zone connection between two stations. An example of a graph representing a macroscopic railway network can be found in Figure 2.3.

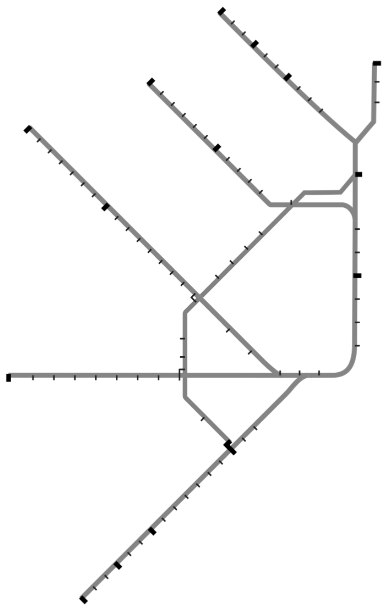
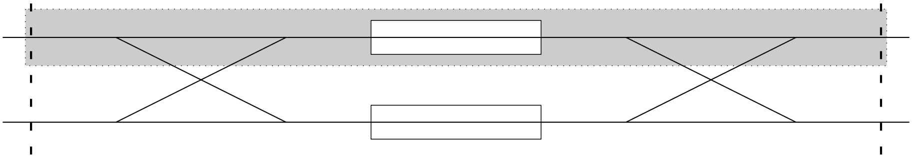
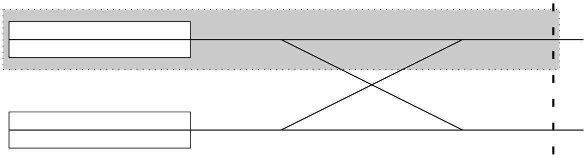


Figure 2.3: Representation of a macroscopic railway network (Burggraeve et al., 2017). The black indications represent stations. The bold black indications represent terminal stations. The gray lines indicate a connection between two stations.

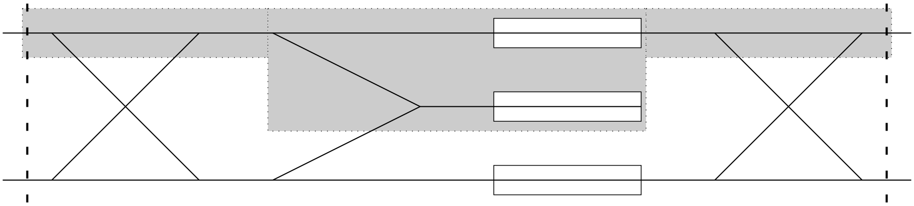
The terminology of *station area* is used in two ways in this dissertation. In Chapter 3 and 4, it defines a group of stations very close to each other and connected with grid zones. A station area can be divided in *platform areas* (one platform area for each station) and grid zones in between these platform areas. So here a station area coincides more or less with an area that contains several stations close to each other. In Chapter 5, however, a station area coincides with the area containing and surrounding one station. The network that is considered in Chapter 5 has a rather simple structure. A railway network is



(a) Typical intermediate non-terminal station where one of the station areas is indicated by the colored rectangle. A train enters that station area if it enters the colored rectangle.



(b) Terminal station with two platforms, which can both be used for turning around. One of the station areas is indicated by the colored rectangle. A train coming from top right enters the station area if it enters the colored rectangle.



(c) A station with one extra platform, referred to as an intermediate terminal station. This platform is only connected with the tracks at one side and can be used for turning. For a train arriving from top left, the station area is indicated by the colored rectangles. This train enters the station area if it enters the most left colored rectangle.

Figure 2.4: Three station types in a simple network. The vertical dashed lines situate the location of the signals before and behind the station. The white rectangles represent the platforms. The crosses at both sides of the platforms represent the switches and tracks that connect both station areas.

defined to be *simple* if (i) in between two succeeding stations, there is one track in each direction, (ii) in each station there is one platform in each direction, (iii) in each intermediate terminal station (see Figure 2.4c), there is one extra platform for turning, (iv) the 'assembling' of tracks coming from different terminal stations occurs within stations. Everywhere outside the stations in a simple network there are bridges and tunnels to avoid the crossing of tracks.

Moreover, overtaking is not allowed in a simple network. A station area in a simple network has to be interpreted as follows. It consists of the switches just before and behind the station and the platform belonging to one direction to go through the station. The DSB S-tog network of Copenhagen and many metro networks are examples of a simple railway network. For a visual representation of the different station types in a simple railway network, see Figure 2.4. Note that a station in a simple network consists of two station areas, one in each direction, according to the interpretation of station area used in Chapter 5. This is illustrated in Figure 2.4a.

2.1.2 How does a train drive through a railway network?

Nowadays, in many railway systems around the world, and in both railway systems considered in this dissertation, trains drive from signal to signal. The main information that can be read from a signal is the following. (i) A train can enter the next section at the maximum speed of that section. In Belgium the signal then shows a green light. (ii) A train has to stop before the signal. The train may not enter its next section, since there is another train or an obstacle using (a part of) this section. In Belgium, the signal then shows a red light. (iii) A train can enter the next section while adhering to a speed reduction. In Belgium, the signal then shows two yellow lights, i.e. double yellow. This signal then indicates that there is no train in the next section, but there is a train or obstacle in the section following the next section. This is why the speed reduction is necessary. The train should be able to stop before the next signal if necessary.

In Belgium, if a train passes a green or a double yellow signal, this signal turns red to warn the next approaching train. Just as is required that only one train at a time can use a platform, this condition also holds for other parts of the network. A bit before the train passes a green or a double yellow signal, the section is exclusively allocated to this train. No other train may then enter this section. An example is provided in Figure 2.5.

A section is exclusively allocated to a train during the following time intervals (Hansen and Pachl, 2008):

- the time for clearing the signal;
- the time for watching the approach indication;
- the time for approaching the signal at the entrance of the section;
- the time for driving through the section;

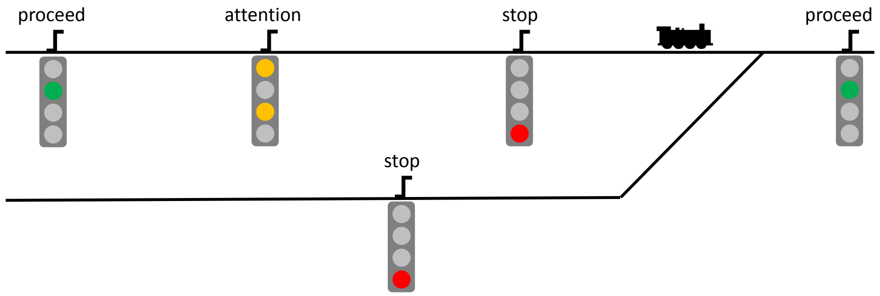


Figure 2.5: Information provided by signals (in Belgium).

- the time for clearing the section;
- the time for unlocking the section.

A section that is exclusively allocated to a train is said to be blocked by that train. The *blocking time* is the total time a section is allocated exclusively to a train and therefore blocked for other trains. A section can contain (part of) links and zero up to several nodes. Since a section is always blocked as a whole, a section is also referred to as a *block section*. A train *reserves* a section and the nodes in this section at the instant that this section is allocated exclusively to this train. This time instant is referred to as the *reservation time*. The train *releases* a section and all infrastructure elements in this section at the end of the blocking time. This time instant is referred to as the *release time*.

If the microscopic network information is not known, the exact blocking times cannot be calculated. Sometimes the microscopic network information is too extensive, such that the amount of information is first reduced. From the reduced information, however, the exact blocking times cannot be calculated anymore. Then, typically, the time that a part of the network is exclusively allocated to a train is simplified to the time that the train actually drives through that part together with some normative amount of extra time to assure safety. This extra time is determined in such a way that it allows an allocation of this network part to the next train immediately after the first train releases it. It guarantees that the first train is already sufficiently far away when another train wants to reserve the infrastructure, e.g. a station area. The allocated time interval is in this case referred to as the *occupation interval*. The *occupation time* is the length of the occupation interval.

In this dissertation, blocking times are used in timetabling on the signaling level, which will later be defined as microscopic timetabling, (Chapter 3 and 4), and occupation intervals are used in timetabling on a less detailed level, which will later be defined as macroscopic timetabling, (Chapter 5). In the latter chapter, the terminology of reservation and release times is used to indicate the begin and end time of the occupation interval. The *entry time* is the time when the head of the train actually enters the allocated network part and the *exit time* is when the tail of the train leaves this network part. In Chapter 5, the reservation time is always a fixed amount of time before the train enters the station area, independently of the station area. A train enters a station area when it passes the vertical dashed line in Figure 2.4 and enters the colored rectangle.

In case a train stops in a station, the time for driving through the station also contains the time for dwelling at the platform, both at the microscopic and at the macroscopic level. This time is called the *dwell time*. In Chapter 5, the occupation interval of trains not dwelling in a station area is artificially lengthened such that their occupation time is equal to the occupation time of dwelling trains. This decision is made to avoid overtaking in the simple network.

A *route* contains the information on which sections a train will have to pass in the network. It can be described by a sequence of signals or a sequence of succeeding nodes and links. A route can be visualized in a microscopic network graph, see Figure 2.6.

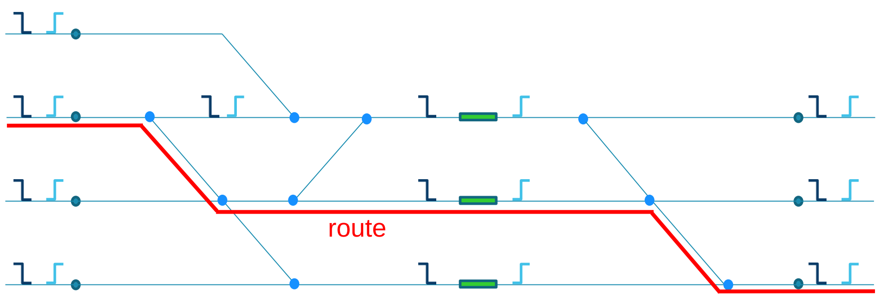


Figure 2.6: Representation of a route.

A *line* is often taken to be a route in a macroscopic network graph ignoring precise details of platforms, junctions, etc. In this dissertation, a line is a sequence of stations in the network together with a frequency and a stopping pattern at these stations. A line may either stop at or bypass a station (which saves time for bypassing passengers). A line is visualized in Figure 2.7.

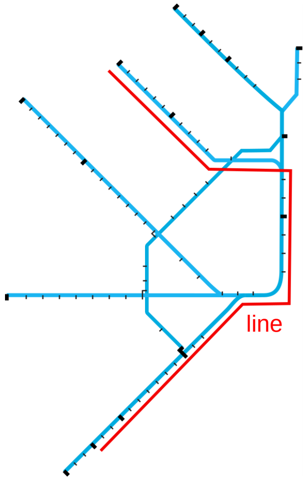


Figure 2.7: Representation of a line.

Based on a train route one can determine the blocking times of the sections on this route. If only a train line is known, then estimations can be made for the time that the train is in the network. Therefore the following terminology is used in Chapter 5.

The *running time* between two succeeding station areas is the time that a train needs between the release time of the first station area and the reservation time of the next station area.

The *drive time* between two succeeding stations is defined as the occupation time of the first station area and the running time to arrive at the next station area, so it is the time that a train needs between the reservation time of the first station area and the reservation time of the next station area. Since the reservation time of a station area is defined as a fixed amount of time before the entry time of that station area in Chapter 5, the drive time between two succeeding station areas also coincides with the time between the entry times in these two station areas. A visual representation of the different kinds of train travel time concepts is provided in Figure 2.8.

The *necessary turn time* is the time for the train to enter the station area of a terminal station (decreasing speed), stopping at the platform, alighting and boarding of passengers, extra time needed by the driver to move from one side of the train to the other side and the time for the train to leave the station area of the terminal station again (increasing speed). The necessary turn time is in

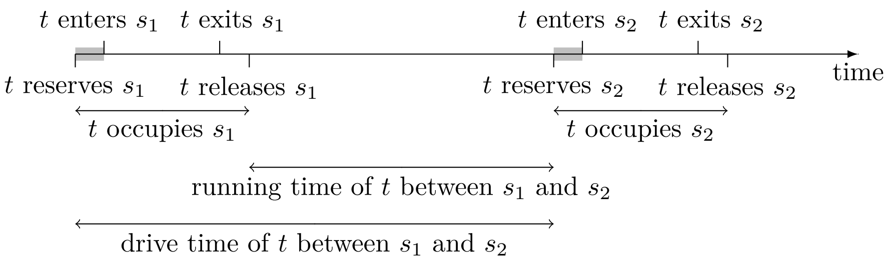


Figure 2.8: Representation of the reservation time, the entry time, the release time, the exit time, the occupation time, the running time and the drive time of train t for two succeeding stations s_1 and s_2 as used in Chapter 5. The parts indicated in gray are equally long, independent of the involved train and stations.

fact the shortest possible occupation time of a train in a terminal station.

2.1.3 How to plan a railway system?

A railway system is typically planned in several steps that are dependent on the planning horizon (Huisman et al., 2005). An overview of these railway planning steps is visualized in Figure 2.9 (Lusby et al., 2011). Furthermore, there are some more planning steps that are not visualized in Figure 2.9, because they are not strictly necessary, sometimes included in other planning steps or not easily categorized based on a time horizon. Examples are maintenance planning and passenger routing.

Figure 2.9 shows the planning steps in the order that they are usually run through in practice. At the *strategic level*, long term decisions are made, at the *tactical level*, medium term decisions are made and at the *operational level*, real-time decisions are made. This dissertation presents developments in line planning, railway routing and timetabling. So, the rest of this section focuses on the considered problems, input, output and other considerations for line planning, timetabling and railway routing. Decisions on building new infrastructure or removing superfluous infrastructure are determined in *network planning*. *Crew scheduling* is the assignment of the crew to the work rosters, where rest, preferences, place of residence, etc. are taken into account. *Rolling stock scheduling* is the assignment of rolling stock to the planned lines. Here, passenger demand, platform lengths, infrastructure restrictions, etc. have to be taken into account. Solving disturbances and delays during daily operation are covered by *real time dispatching*.

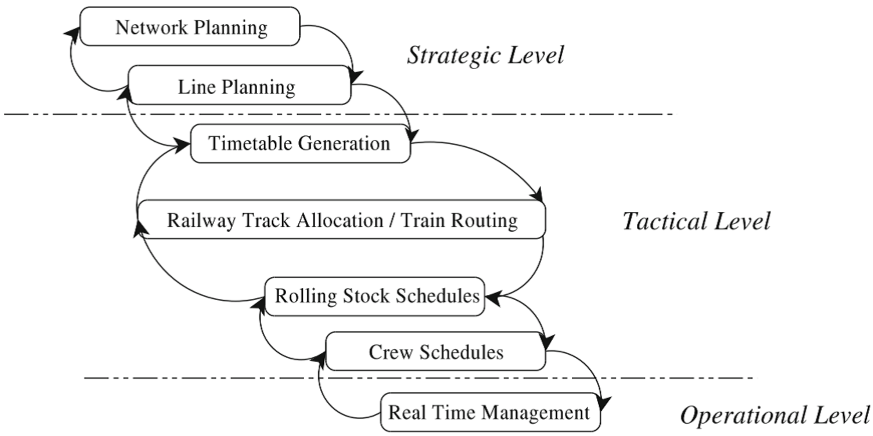


Figure 2.9: Railway planning steps (Lusby et al., 2011).

In the current literature on railway planning, sometimes, up to three planning steps are solved together, e.g. Michaelis and Schöbel (2009, line planning, timetabling and rolling stock (vehicle scheduling)), Bach et al. (2016, timetabling, crew scheduling and rolling stock (engine scheduling)), Schmidt and Schöbel (2015, line planning, timetabling and passenger routing), Gattermann et al. (2016, line planning, timetabling and passenger routing), Lidén and Joborn (2017, timetabling, railway routing and maintenance planning), but mostly only one planning step is considered at a time. Each of the separate planning problems is already difficult on its own, even NP hard according to (Schöbel, 2015). Unfortunately, a solution of a previous planning step does not necessarily allow a good, or even a feasible, solution for the next planning step. Decisions at each of these planning steps affect the performance of the decisions at the other planning steps. Therefore, the arrows in Figure 2.9 indicate the necessary feedback between the planning steps.

Line planning

Railway line planning is the problem of constructing the set of lines that will be operated in the railway network. The input is typically the infrastructure lay-out and an origin-destination matrix of passenger demand. The output is a *line plan* that consists of a set of lines each with a stopping pattern and a frequency. The following considerations can be taken into account while making a line plan.

- The line plan must assure a *minimum service* at every station. This means that per time period at least a certain number of trains has to make a stop at every station. This minimum number of trains typically depends on the station characteristics and the passenger demand.
- The line plan must satisfy a *budget constraint*. In that case an operational cost is determined for each line and the available budget is specified.
- Only a subset of all lines that can be created are considered, for example based on stopping pattern, length or frequency. The set of allowed lines, i.e. lines that are considered in the line planning, is called the *line pool*.
- The line plan must satisfy the *network capacity constraints*. Only a limited amount of trains can traverse a station or use the network in a limited amount of time.
- The line plan must assure a certain percentage of *direct connections* to the passengers. If a line connects two stations where it makes a stop, then there is a direct connection between these two stations.

This list is indicative to introduce the italicized concepts in line planning. This list is not exhaustive covering all considerations.

Whilst most railway planning steps cannot create an output without considering a solution of the previous planning step, this is not true for the routing and the timetabling. Both planning steps can be interchanged. So, even though timetabling is presented above railway routing in Figure 2.9, here, railway routing is discussed first. The reason is that the availability of a routing plan determines if microscopic or only macroscopic timetabling can be done. In order to explain the difference between both timetabling problems, it is useful to first introduce the routing problem.

Railway routing

Railway routing is the problem of assigning trains to routes. The output is a *routing plan* which determines for each train exactly which parts of the infrastructure it will use. Typically, railway routing is only worthwhile studying in a station area or in a grid zone. The line plan is considered as input and it determines where the trains enter and leave the station area or grid zone. The routing plan determines which specific nodes and sections the trains use. In large stations or grid zones, many alternative route options are available for each train.

As mentioned above, the construction and optimization of a routing plan and a timetable are closely interwoven. The order in which both problems are solved, determines the complexity, the restrictions and the objectives of both problems. In case the timetabling problem is solved first, only a macroscopic timetable can be constructed. The routing problem, which is solved thereafter, is then constrained by the macroscopic timetable. In case a macroscopic timetable is available, the assignment of the routing plan incurs at the same time a microscopic timetable. Thus, in this case, the routing plan is only feasible if it incurs a conflict-free microscopic timetable. Not every macroscopic timetable, however, does assure this existence of a feasible routing plan (Sels, 2016b). Thus, only after solving the routing problem, a statement can be made about the conflict-freeness of the schedule. In case the routing problem is solved first, the routing plan is not constrained by the timetable. Once the routing plan is known, not only a macroscopic timetable, but immediately a microscopic timetable can be designed. In this case the timetable is constrained by the routing plan. A conflict-free timetable, however, does not exist for every routing plan. Only after solving the timetabling problem, a statement can be made about the conflict-freeness of the schedule. Thus, both a routing plan and a timetable are necessary to judge the conflict-freeness of the schedule.

For large networks with many stations but relatively simple infrastructure layout or relatively sparse traffic, it is advantageous to first construct a macroscopic timetable and only thereafter consider railway routing. In a railway bottleneck, by contrast, it could be advantageous to immediately look at the microscopic infrastructure level for the construction of an optimal routing plan and timetable. This could lead to a more efficient use of the available infrastructure, which is illustrated in Chapter 4.

The following considerations can be taken into account while making the routing plan.

- Only a subset of all possible routes that can be created are considered, based on the place where the train enters the station area and the type of the train or based on the track directions imposed by the railway infrastructure manager.
- In case timetabling is done first, the construction of a routing plan is restricted by the timetable. In case railway routing is done first, the routing plan restricts the timetabling.
- As is the case for a line plan, the routing plan must satisfy the capacity constraints of the network, i.e. the capacity constraints of sections, tracks and nodes. Only a limited amount of trains can traverse a section or use a track or node in a limited amount of time.

In the next subsection, we go deeper into the timetabling and routing restrictions dependent on which problem is solved first.

Timetabling

Timetabling is the problem of assigning precise utilization times for infrastructure resources to every train in the rail system. First, a short overview is given of how timetabling is currently done in practice in Belgium. Then this subsection is split into a part on microscopic timetabling and macroscopic timetabling.

On the Belgian railway network, there is one main operator for all national passenger trains, NMBS. Furthermore, there are some international train operators (e.g. TGV, Eurostar), but they only cover a small percentage of the passenger trains in Belgium. Finally, there are also national and international freight trains. Since freight trains generally do not pass through railway bottlenecks, they are not taken into account in this dissertation. Concerning the national passenger trains, the operator NMBS ultimately determines the timetable. There are two main phases in the timetabling process: constructing the timetable and adapting the timetable based on an evaluation of the rolling stock assignment. For constructing the timetable, NMBS will initially submit requests for train lines and desired arrival and departure times based on the data from ticket and railcard sales and marketing information (from passenger organizations, schools, ...). Infrabel, the Belgian railway infrastructure manager, will make a draft offer in which it also takes the following requirements of NMBS into account: trains with similar trips roughly have to be equally spread over the period of the timetable, the timetable must be symmetric and certain passenger transfers must be assured with connecting trains preferably sharing the same platform. Before proposing a draft offer to NMBS, Infrabel compares the performance of different offers by using simulation. Infrabel then proposes a draft offer with an acceptable balance between allocated capacity and robustness to NMBS. This draft offer is the basis for alternative proposals in an iterative process, until the final offer is accepted by NMBS. In these iterations, the second phase in which the timetable is adapted based on an evaluation of the rolling stock assignment also takes place: rolling stock reuse cycles are fixed in the terminal stations and maintenance and storage in shunt yards are planned. For international trains, there are agreements with the other involved countries on the preferred times that these trains cross the border. These international trains get priority and the resulting restrictions on arrival and departure times are taken into account by Infrabel when making and checking the draft offers for the national trains. The routing of national, international, passenger and freight trains through the Belgian stations is planned by Infrabel, but is sometimes adjusted in real-time due to disturbances.

Microscopic timetabling is the problem of assigning blocking times to sections and nodes on the trains' routes. To construct a microscopic timetable, a routing plan must be known. The assigned times must ensure that trains can follow their routes in the network, stop at appropriate stations where necessary, and avoid any conflicts with other trains. A *conflict* arises where two trains want to use the same part of the infrastructure at the same time, for example at a section, switch, platform or turning track. According to Caimi (2009) the assignment of reservation and release times is conflict-free, if no two trains (are planned to) block the same infrastructure element at the same time. According to Bešinović et al. (2016a) a timetable is feasible if all trains are able to adhere to the schedule on their assigned routes, in their own words (cited): *"if (i) the individual processes are realizable within their scheduled process times, and (ii) the scheduled train paths are conflict free, i.e., all trains can proceed undisturbed by other traffic."*

Macroscopic timetabling is the problem of assigning arrival and departure times to the stations on the trains' lines. Here a routing plan is not strictly needed. It is easier to make a macroscopic timetable, but it is more difficult to identify a conflict in macroscopic timetabling. Normative macroscopic times are used in the calculations to avoid conflicts and assure safety. These normative macroscopic times can be seen as norms (e.g. for headways and occupation times) that must be satisfied to assure a macroscopic feasible/conflict-free timetable. Approximating accurate occupation times and minimum headway times with these normative macroscopic times can, however, not assure conflict-freeness at the microscopic level. *Headway times* can be defined as the necessary safety time between arrivals and departure of trains in stations or between occupation times at other parts of the network. Without checking at the microscopic scale, a macroscopic timetable can never be assured to be microscopically conflict-free.

In Chapter 3 and 4 microscopic timetables are constructed. In Chapter 5, macroscopic timetables are constructed. In practice, macroscopic timetables are presented to the passengers (when microscopically conflict-free), while microscopic timetables are used by train drivers and dispatchers.

A timetable can be represented in a time-distance diagram. Figure 2.10 is a first example. This representation is more appropriate for visualizing macroscopic timetables. Stations and sections between stations are represented on the vertical axis and time on the horizontal axis. A horizontal part of a time-distance path represents the occupation interval in a station where the train has a stop.

Figure 2.11 is a second example. This representation is more appropriate for visualizing microscopic timetables. Here, sections are represented on the horizontal axis and time on the vertical axis. The presence of blocks can be explained by the fact that a train always blocks a whole section. The blocks

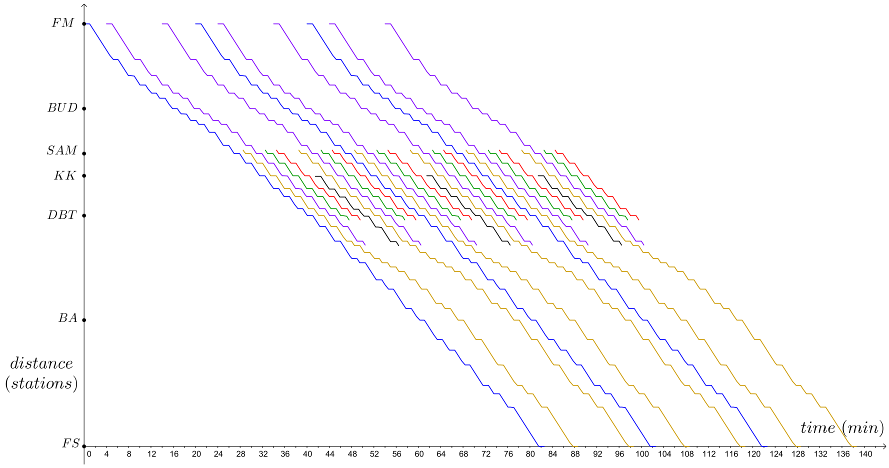


Figure 2.10: Time-distance diagram.

overlap in time because of the inclusion of some safety time before the train actually enters a section and the time that the head of the train is already in the next section, while its tail is still in the previous section.

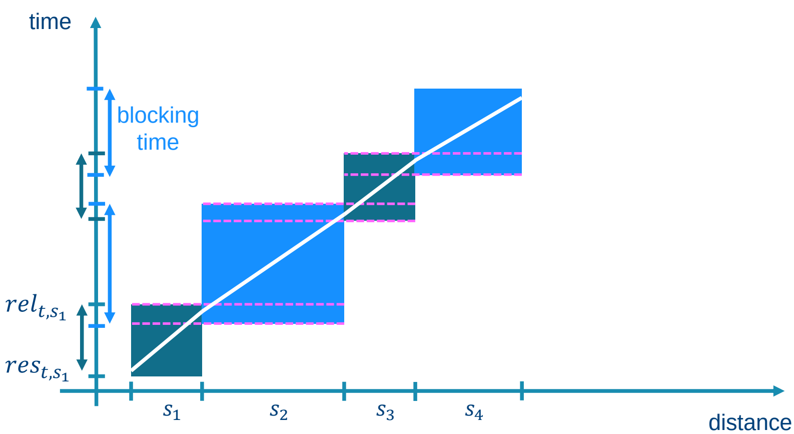


Figure 2.11: Blocking time diagram of one train blocking sections s_1, s_2, s_3 and s_4 consecutively. The pink dotted lines indicate the time overlap of the blocking times of subsequent sections.

Figure 2.12 is a third example. This representation is useful to present

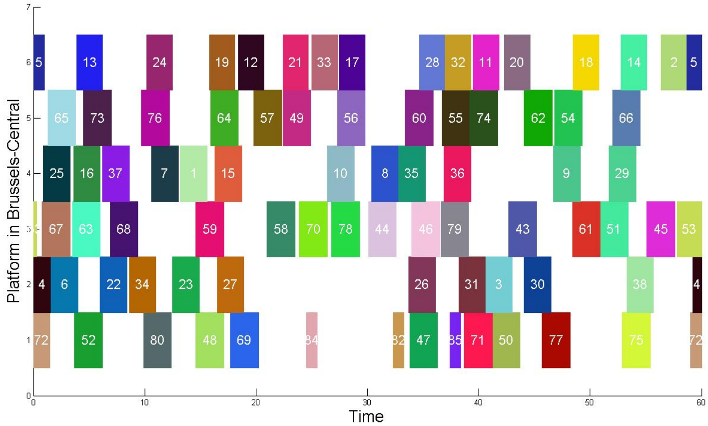


Figure 2.12: Blocking time diagram for the six platforms in Brussels-Central. These nodes are blocked by 11 up to 15 trains during the period of one hour (60 minutes).

microscopic timetables on networks with a complex infrastructure lay-out, for example in a grid zone or a station with many platforms. The disadvantage of the time-distance diagram in Figure 2.11 is that two trains of which the routes only share one node or a part of a section cannot be visualized together except on the shared part. In Figure 2.12, the blocking times for each node are visualized. The nodes are listed on the vertical axis and the time is represented on the horizontal axis. The advantage is that it can be easily observed when and by which train a node is blocked. The disadvantage is that it is not possible to follow the trajectory of one single train.

A timetable will not be efficient if train drivers have to stop at signals outside the platform area. Even if such an inefficient stop is not planned, it may occur in practice when trains are scheduled too tightly. Train delays can be categorized in two categories based on their cause. If a train gets delayed because of an external disturbance, e.g. significantly more passengers than expected are boarding the train, material failure, snow, copper theft, etc., its delay is called a *primary delay*. If a train gets delayed because of another train that is delayed, its delay is called a *secondary delay* or a *knock-on delay*. The following considerations may be taken into account while making a timetable.

- Buffer times may be included in the timetable. The *buffer interval* between two trains on a part of the network (e.g. node, section, switch, turning track, etc.) is the time interval between the release time of the first train of that part of the network and the reservation time of the second train of that same part. In other words, the *buffer interval* is the time interval between the blocking times of these two trains on that part of the network. The *buffer time* is the length of the buffer interval and thus the time between the blocking times of two trains on a part of the network. Buffer times are useful to avoid propagation of delays between trains, i.e. to avoid that a delayed train affects a next train using the same part of the network. Buffer times do not lengthen the train travel times, but they can lengthen the passenger travel times in case of transfers. In Chapter 3 and 4, the buffer time between the reservation and release times of nodes is optimized while in Chapter 5, the focus is on the optimization of the buffer times between the reservation and release times of station areas. A feasible timetable may contain a buffer time equal to zero. In that case the blocking times are still respected for a microscopic feasible timetable and the safety headways are respected for a macroscopic feasible timetable. The infrastructure is then allocated to two succeeding trains without a time gap in between. A *safety headway* is the necessary time, i.e. a normative amount of time fixed in advance, between two events (e.g. arrival, departure, reservation, release, etc.) to assure safety. A headway is in fact an estimation of the blocking time to be used at macroscopic level. So headways serve the same goal as blocking times. Normally, these safety headways are fixed on the safe side, but microscopic feasibility can never be assumed until it is proven.

- Supplements may be included in the timetable. A *supplement* is an extra amount of time that is added to the necessary drive or running time of a single train on a part of the network. Supplements are useful for trains to absorb their own delay during operation. A disadvantage is that supplements increase the planned travel times of trains and passengers.

- The timetable can be cyclic or non-cyclic. A *cyclic timetable* is a timetable that repeats itself every period, e.g. every hour. This is not the case for a non-cyclic timetable. In many European countries cyclic train timetables are used, which makes the timetable easier to remember for the passengers. The timetables constructed in this dissertation are always cyclic.

- The timetable can be symmetric or asymmetric. A *symmetric timetable* is a cyclic timetable where trains driving the same trajectory in opposite direction pass the stations on their route at timings that sum up to a multiple of the period length. This assures that transfer times between trains are equally long in both directions, i.e. the transfer time between trains A and B equals the transfer time between trains B^{opp} and A^{opp}, where A^{opp} (B^{opp}) indicates the opposite train of train A (B). In a symmetric timetable, the time-distance graphs of each pair of opposite trains are mirrored around the same symmetry-axis. More details on symmetric timetables can be found in (Liebchen, 2003).

Note that the more supplements and buffer times are imposed on timetabling, the less freedom there is to find a feasible conflict-free schedule.

2.2 Robustness in railway planning

In diverse disciplines, a robust schedule is typically interpreted as a schedule that is unaffected and equally good performing in case of disturbances and delays. In the rest of this dissertation, robustness will refer to this typical/traditional interpretation. In a railway context, if a train gets delayed, a robust schedule is typically interpreted as a schedule that assures that passengers still arrive on time at their destination and that no other trains are affected by the delay (Cicerone et al., 2009, Cacchiani and Toth, 2012, Bešinović et al., 2016a). It is easy to see however, that e.g. if the train from Leuven to Brussels would be planned to take two hours for this trajectory, which currently takes only 18 minutes, then this train will almost never arrive too late. However, this is not really what passengers prefer. So eliminating delays is not the only objective for delivering a good service. Passengers prefer both short and reliable travel times. Here reliable travel times refer to the travel times announced by the planning. Short and reliable travel times seem counteracting objectives, however, minimizing ‘the passenger travel time in practice’ assures both aspects. Minimizing the passenger travel time in practice means that the passenger travel time, including transfers, is minimized while taking frequently occurring small delays into account. So a *passenger robust railway system* is a railway system that minimizes the total passenger travel time in practice in case of frequently occurring small delays. If a passenger robust schedule is implemented, the travel time is still reliable in case of a (small) delay. A schedule is called more passenger robust than another schedule if its total travel time in practice is less. Parbo et al. (2016) give an extensive overview of passenger perspectives in railway timetabling.

Different planning steps influence the passenger travel time in practice. For example, line planning decides on direct connections and the choice between local and intercity trains, which affect the travel time between origins and destinations. The routing plan (including platform assignment) decides on which trains can directly affect each other and thus propagate delays to each other and make travel times less reliable. The timetable decides on the arrival and departure times in the stations and thus also fixes the transfer times for the passengers. The real time dispatching decides on how a conflict is solved. This also has a direct effect on the passenger travel time in practice. Furthermore, to compute the exact passenger travel time in practice, decisions on these different planning levels must be known together with the frequently occurring small delays in the network and the size of each passenger flow through the network. In this dissertation the distribution of frequently occurring small delays is assumed to be known, i.e. they are based on historical delays, and the dispatching strategy is simplified to a first come first serve approach to solve real time conflicts.

The rest of this section presents an overview on how (passenger) robustness is implemented in line planning, timetabling and railway routing. The overview is restricted to these planning levels, since these are the planning levels for which developments are discussed further in this dissertation. This overview does not focus on the different interpretations of robustness but on how robustness can be obtained across different planning levels. A review on the different interpretations of robustness can be found in Dewilde et al. (2011) and Andersson (2014). The book 'Robust and Online Large-Scale Optimization' (Ahuja et al., 2009) assembles worthwhile research in which robustness is implemented in railway planning. A very recent review on how robustness can be obtained, can be found in (Lusby et al., 2017).

Lusby et al. (2017) categorize the literature based on their target planning level. In order for this dissertation to be complementary, the literature on the implementation of robustness will be divided into two classes. A first class focuses on the evaluation of (passenger) robustness. Evaluation methods found in literature are either based on robustness indices or based on simulation. The second class includes (passenger) robustness in the optimization process. Optimization methods found in literature either use a robustness index as an optimization criterion or they take several delay scenarios into account, these latter methods can be classified as stochastic optimization or as robust optimization.

Evaluation of (passenger) robustness by a robustness index

A *robustness index* is a quantitative rating derived from some analytical properties of the schedule. The advantage is that such a rating can typically be calculated without large computational efforts. The disadvantage is that it only gives a limited amount of information. There exist robustness indices that can be calculated from both macroscopic and microscopic schedules, but there also exist robustness indices that can only be derived from microscopic schedules.

A high variety of robustness indices are present in the literature. Some of the more straightforward indices are the minimal buffer time, the amount of supplements, the maximum number of trains that make use of the same infrastructure, the blocking time diagrams of busy stations (e.g. Dewilde et al., 2011, Salido et al., 2012, Jensen and Landex, 2013, Jovanović et al., 2017). Some of the more sophisticated robustness indices and how they can be obtained will now be discussed.

Fischetti et al. (2009) use a linear programming validation tool to measure robustness. This tool makes use of the assumption that a robust timetable has to favor delay compensation without heavy human action, such that no complex dispatching strategies need to be implemented. As a result, train cancellations and changing the order of trains are not allowed. Small adjustments made to the timetable are used as input. The model minimizes the necessary adaptations to the original timetable necessary to satisfy the adjusted travel, dwell or headway times. The objective value of the linear programming model is their robustness index, which can be used to compare timetables (for the same line plan).

De-Los-Santos et al. (2011) propose several robustness indices to evaluate a network and a line plan at the same time. These indices are based on the total travel time in the network when a link fails, depending on the origin-destination matrix. The authors say that a network is robust if it maintains its basic functionality under the failure of some of its components. In (Laporte et al., 2010) one of these indices is applied to design a railway transit network.

Peterson (2012) looks for robustness measures that can be taken for single services, without disturbing other traffic. The author obtains these measures by analyzing the punctuality and its variance and how both evolve en route. He compares the results to schedules where the available margin in the schedule is rescheduled proportionally, evenly distributed over time, or according to demand.

Andersson et al. (2013) introduce critical points as a robustness index. Cited from (Andersson et al., 2013): *"Critical points refer to very time-sensitive dependencies between different pairs of trains at different locations in the network. In the context, which this paper is focused on, such points typically occur when trains enter a line behind an already operating train, or where trains overtake each other."* These points can give planners specific insight in how to improve the schedule by increasing the running time or the headway margin for the trains passing this point. These points can also be used to compare the quality of different timetables.

Sels et al. (2015) use an analytical function as the robustness index to evaluate a macroscopic timetable. They calculate the expected time of all actions (ride, dwell, transfer, knock-on delay) separately and add them up to estimate the expected passenger travel time. Therefore they use an analytic expression in function of the minimum time for each action and the assigned time supplement to an action. This function takes for each action a delay distribution for primary delays into account. The underlying assumption is that the primary delays of the different actions are independent of each other.

Evaluation of (passenger) robustness by simulation

Another way to evaluate (passenger) robustness is to use simulation. This approach is more cumbersome than using a robustness index, but it has the great advantage that the performance of the schedule in practice can be realistically forecasted. One type of simulation is a Monte Carlo discrete event simulation. Here trains enter the network with a delay drawn from a delay distribution or trains encounter such a delay while driving through the network, for example while dwelling in a(n overcrowded) station. The delay distribution is assumed to be a good approximation of delays that will be encountered in practice. This delay distribution can be based on historical delays as is the case in (Dewilde et al., 2013, 2014a) and (Dewilde, 2014b). Often an exponential distribution is assumed (Goverde, 1998). In order to get a realistic image of the performance of the schedule, a high number of delays is drawn from this distribution (Vansteenwegen, 2008, Dewilde et al., 2013, 2014a), for example up to 10 000 draws. Also in this dissertation, a high number of delays is drawn from an exponential distribution, based on historical delays, to simulate the schedule. For each train, a timing is associated with its next action. The time always jumps to the next timing associated with a train action, which explains the terminology of 'discrete event'. At the timing of a train action, this action is first checked for feasibility and executed if feasible, otherwise the train action is postponed till a later timing. In both cases, a new timing is associated with this train. A postponement can correspond to a (knock-on) delay for this train or the use of a supplement. In case multiple train actions are planned at the same time, a priority rule is necessary to decide which action to execute first.

A continuous simulation is another type of simulation. The difference with discrete event simulation is that the state of the trains changes continuously over time and not only when they block a next section or get delayed. This is used in real time simulations, like in the railway planning software OPEN TRACK that is used by many railway organizations all over the world.[3]

Both discrete event and continuous simulation include a dispatching strategy to solve conflicts, i.e. a priority rule when two trains want to block the same

[3]OPEN TRACK: http://www.opentrack.ch/

infrastructure at the same moment. This dispatching strategy affects the travel times of the passengers in practice. The validation of the schedule thus also depends on the dispatching strategy that is implemented. No best dispatching strategy is known and confirmed yet. In practice, real time management is currently based on the experience of the dispatchers and in research, real time management is a major research topic on its own. The simplest dispatching strategy is '*first come first serve*', where the first train that arrives always gets priority. The advantage is its simplicity, a disadvantage is that it can result in, for instance, a fast train ending up behind a slow train.

Different measures exist to interpret the output of a simulation and to judge whether one schedule is more (passenger) robust than another schedule. In case of passenger robustness, one looks at the average total travel time of all passengers over all simulation runs (each simulation run with a different draw from the delay distribution) (Dewilde et al., 2013, 2014a). Other interpretations of robustness used in literature are the average amount of knock-on delay (Dewilde et al., 2013, 2014a), the average percentage of delayed trains (Dewilde et al., 2013, 2014a), the percentage of missed connections (Vansteenwegen, 2008), the percentage of transfer times longer than 20 minutes (Vansteenwegen, 2008), the total stopping cost (Vansteenwegen, 2008), the passenger disutility (Takeuchi et al., 2007), the absorption of delays (Bešinović et al., 2016a), etc. These measures are also often referred to as robustness indices.

(Passenger) robustness as railway optimization criterion

Several different objective functions are used to include (passenger) robustness in railway optimization. Robustness indices that are earlier discussed and that are used to evaluate the robustness of a railway schedule, can usually also be used as optimization criterion, (e.g. Sels et al., 2015, 2016a, De-Los-Santos et al., 2011).

Often (passenger) robustness is only indirectly optimized by choosing an optimization criterion that is correlated to (passenger) robustness. If this criterion improves, the (passenger) robustness is assumed to improve. The advantage is that it is often easier to optimize this alternative criterion. The disadvantage is that the alternative criterion does not give a complete (one-to-one) coverage. Close to optimality, for example, the alternative criterion can still be improved, while the (passenger) robustness already starts deteriorating, or the alternative criterion is proven to be optimal, while the (passenger) robustness can still be further improved.

In the matheuristic of Dewilde et al. (2013, 2014a), the minimization of the inverse buffer times is used as robustness criterion. The idea here is that if the buffer times between the trains are large enough, then delayed trains will not propagate their delays to other trains. Only the travel time in practice of the passengers on the initially delayed train is lengthened. In this dissertation, this same robustness criterion is used in the objective function, but the inverse buffer times are weighted based on passenger numbers and recurring delays of the trains (Chapter 3). These weights have the advantage that buffer times before crowded trains and buffer times after often delayed trains are prioritized. Furthermore, in this dissertation also an altered version of this criterion is used by maximizing the minimum buffer time over all trains (Chapter 4 and 5). Minimizing the inverse buffer times makes it more advantageous to increase smaller buffer times than larger buffer times. However, minimizing the inverse buffer times also makes the objective function non-linear.

Note that if the routing plan, the supplements and the running and dwell times are fixed, the total amount of buffer time is the same for all timetables (for the same line plan), only the spreading of the buffer times differs. By maximizing the ‘minimum’ buffer time the spreading between the trains becomes more equal, which globally lowers the risk to propagate delays.

Apart from passenger robustness as objective in railway timetabling, another objective with a focus on passengers, is the minimization of the average passenger delay. This average passenger delay can differ from the average train delay to a large extent, as trains do not transport the same number of passengers and because a passenger can need multiple trains for his trip such that he/she can be delayed as a consequence of a missed transfer. Other research that focuses on passengers is for example described in (Vansteenwegen and Van Oudheusden, 2006, Sels et al., 2013, Engelhardt-Funke and Kolonko, 2004, Takeuchi et al., 2007, Liebchen et al., 2010).

The work of Vansteenwegen and Van Oudheusden (2006) improves passenger robustness by only focusing on transfers. In Vansteenwegen (2008) the time that passengers spend waiting is optimized by changing the amount, location and or distribution of supplements. This waiting time includes waiting time in stations, deviations from ideal supplements and extended transfer times. Also Engelhardt-Funke and Kolonko (2004) concentrate on the passenger’s waiting time, but take at the same time investment costs into account and look for a Pareto optimum.

In (Cicerone et al., 2009) robustness is aimed for by an algorithm that minimizes the overall delay faced by the total passenger population by proportionally or additively adding slack time in the schedule. They strive for *recoverable robustness*, i.e. they look for a resilient schedule that is able to absorb small delays by possibly applying given limited recovery capabilities.

Liebchen et al. (2010) improve passenger service by increasing the nominal travel time, so the planned travel time. Including supplements in the planned travel time gives a train the opportunity to absorb its delays and to stick to its schedule.

Bešinović et al. (2016a, 2017) and Goverde et al. (2016) developed an integrated micro-macro approach to construct a microscopic conflict-free timetable. The objective of their macroscopic model minimizes a weighted sum of running, dwell and transfer times, and a robustness cost, which corresponds to the delay settling time in a Monte Carlo simulation.

(Passenger) robustness in robust and stochastic optimization

Robust and stochastic optimization are two general classes of methodologies to handle uncertainty. In both classes delay scenarios are explicitly taken into account during optimization. The set of delay scenarios is fixed beforehand. In robust optimization, it holds that requirements that have to be fulfilled for the default situation, must also be fulfilled for the delay scenarios. In stochastic optimization, simulation of (an) intermediate result(s) is used in the optimization process. Fischetti et al. (2009) make use of a scenario-based stochastic programming approach. This approach couples robust and stochastic optimization. The authors propose an exact optimization model. Their objective is to minimize the cumulative delay for all scenarios and events. The explicit inclusion of the delay scenarios highly increases the computation time of the model, compared to the model that only takes the default case into account. The terminology of *light robustness* and *strict robustness* refers to how strict the requirements for the delay scenarios must be taken into account in robust optimization. Therefore Bertsimas and Sim (2004) defined the *price of robustness*, which shows how much the robust solution differs in objective value concerning the original objective, compared to the optimal solution for this original objective.

A disadvantage of robust optimization is that the exact distribution of the data needs to be known. Another disadvantage is that, for strict robustness, feasibility is required in all scenarios, potentially leading to highly conservative solutions. For stochastic programs, the tractability and efficiency is a challenge. An advantage of robust and stochastic optimization is that the validation of robustness occurs inside the model itself. So in fact, no separate validation step is necessary.

Khan and Zhou (2010) propose a stochastic optimization formulation to allocate slack times in a schedule. They make use of a two-stage recourse model. The first stage consists of the minimization of a utility function of the total trip time and in the second stage the expected delay is minimized. The problem itself is solved with a sequential heuristic approach that iteratively fixes the timetable of one train. It makes use of a stochastic shortest path problem (Miller-Hooks and Mahmassani, 2000). For the second stage a set of random segment running times is generated. There the train travel times and dwell times under each scenario must be determined and the first stage timetable may be updated.

Kroon et al. (2008a) propose a stochastic optimization model to re-allocate the time supplements and the buffer times in a given timetable in order to make the timetable maximally robust against stochastic disturbances. The authors use exponential distributions for the primary disturbances which are assumed to be independent of the timetable. For the optimization, a fixed number of realizations is selected and a random sample of primary disturbances is used.

2.3 Railway planning levels

As mentioned before, in practice, the different planning levels are solved hierarchically. In the case that the output of the previous decision level leads to infeasibility at the next planning step, there are several possible approaches to look for a feasible solution to both planning levels together. First, the outcome of the previous level can be replaced by a second best outcome in the hope that a feasible solution for the next level exists. Secondly, the outcome of the previous level can be specifically oriented towards making a feasible solution for the next level possible, by using case dependent restrictions specifically for this goal. Thirdly, the constraints on the outcome of the next level can be relaxed. These approaches increase the possibility of finding a feasible solution for the next level, but not necessarily guarantee a good outcome for both levels together. Solutions to problems that integrate several planning levels outperform the hierarchical approach (Goerigk et al., 2013, Gattermann et al., 2016), but mostly are computationally demanding. Also in this dissertation it is shown that integrated approaches outperform the hierarchical approach, more specifically for line planning and timetabling and for timetabling and railway routing. In the rest of this chapter, the research presented in this dissertation is compared to the state of the art on railway routing (Section 2.3.1), timetabling (Section 2.3.2), line planning (Section 2.3.3), on integrated approaches for line planning and timetabling (Section 2.3.4) and for timetabling and railway routing (Section 2.3.4).

2.3.1 Railway Routing

Research on railway routing focuses on (complex) station areas and grid zones, since these areas give rise to many possible routes. Crossing and overlapping train routes increase the probability of conflicts. The train platforming problem, i.e. the assignment of trains to platforms e.g. solved by Sels et al. (2014), is in fact also a railway routing problem. An extensive overview of the railway routing problem at station level can be found in the survey paper of Lusby et al. (2011). Sun et al. (2014) give a more recent but short literature overview on routing problems in their paper. Research on railway routing usually uses a macroscopic timetable as input, such that the routing problem is restricted by already scheduled arrival and departure times or bounds. Apart from finding a conflict-free assignment of trains to the routes, also a given objective function is optimized while adhering to (or only slightly modifying) the given macroscopic timetable. Examples of optimization criteria are maximizing the number of trains that can be routed, minimizing the shunting movements, maximizing the preferences of trains for certain platforms or routes, minimizing average travel time, minimizing energy consumption, maximizing user satisfaction, robustness etc. (e.g. Zwaneveld et al., 1996, 2001, Caimi et al., 2011, Sels et al., 2014, Sun et al., 2014).

The railway routing approach proposed in Chapter 4 differs from the existing approaches (e.g. Carey and Lockwood, 1995, Zwaneveld et al., 1996, 2001, Sun et al., 2014), in that it precedes the timetabling problem. Consequently, the construction of the routing plan is not restricted by the timetable, i.e. neither timings nor the order of trains are fixed. This incurs that (all) the trains can be freely spread over the available infrastructure. The objective that is maintained for the routing problem in Chapter 4 is that trains interact with each other as little as possible. This is achieved by cleverly minimizing the node usage over the whole network. Since longer routes and detour routes use more nodes, these are inherently avoided by the objective function. Another application of the routing model proposed in Chapter 4 could be to estimate and assess the spreading of a routing plan constructed for a given timetable. The routing model could then be used to calculate a reference unrestricted spreading of the trains on the available infrastructure.

2.3.2 Timetabling

Timetables can be constructed at the microscopic level and at the macroscopic level. Planning at the macroscopic level only provides arrival and departure times in stations, but does not guarantee that conflict-free train routes through the network can be assigned, for instance, to guide the trains to their platform at the scheduled times. Planning at the microscopic level provides reservation and release times for the sections on the trains' routes and thus requires that the train routes are known. Macroscopic timetabling is discussed first, then the focus shifts to microscopic timetabling.

Serafini and Ukovich (1989)'s seminal paper on the *Periodic Event Scheduling Problem (PESP)* is the foundation of many cyclic macroscopic timetabling models (e.g. Schrijver and Steenbeek, 1993, Nachtigall, 1996, Peeters, 2003, Liebchen, 2006, Liebchen and Möhring, 2007, Kroon et al., 2009, Großmann, 2011, Schmidt and Schöbel, 2015) and is also the framework of the timetabling model proposed in Chapter 5. Alex Schrijver and Adri Steenbeek, together with Erwin Abbink, Pieter-Jan Fioole, Dennis Huisman, Leo Kroon, Roelof Ybema, Bert Meerstadt, Matteo Fischetti and Gábor Maróti, won the Franz Edelman Award in 2008 for their research on a more efficient timetable for the Dutch railways (Kroon et al., 2009). A recent and elaborate discussion on timetabling literature in general and PESP in specific can be found in Sels et al. (2016a). Earlier, Cacchiani and Toth (2012) published a literature review on macroscopic timetabling models, like the PESP and the non-cyclic Train Timetabling Problem (e.g. Carey and Lockwood, 1995, Caprara et al., 2002, Cacchiani et al., 2016).

The PESP model schedules events in a period of the cyclic timetable and takes precedence constraints and relations between events into account. Arrivals and departures of trains at stations or reservations and releases of track sections or station areas are events. If two events are related or can affect each other they form an activity. Examples of activities are the arrival and departure of the same train in a station or the reservation times of a shared switch, platform or station area by two different trains. The PESP model constrains each activity time, which is the time between the two events that define the activity. So, related events are linked to each other by constraints that put an upper bound and/or lower bound on the time duration between these events. The PESP is originally defined without an objective function, but several objective functions for PESP can be found in the literature. In Chapter 5 an objective function that maximizes the (minimum) buffer times between trains using the same part of the infrastructure is added, in order to achieve passenger robustness. The timetabling model in Chapter 5 also takes 'turning', 'providing buffer time' and 'station' activities into account as activities besides the usual running and transfer activities. Furthermore, this model contains extra constraints such that trains of the same line can be equally spread over the period of the cyclic timetable. These constraints coincide with the constraints for the synchronization activities considered in (Siebert and Goerigk, 2013). In that paper the impact of including line frequencies in cyclic timetabling is studied and the authors conclude that it positively and significantly affects the quality of the constructed timetable.

For the non-cyclic timetabling problem, different (exact) models are present in literature. One example uses continuous decision variables for the train departure times at stations and binary decision variables for fixing the order of the trains in stations (Carey and Lockwood, 1995). Preferred departure times are assumed and the objective is to plan the trains as close as possible to these preferred timings. Another example discretizes time and uses binary decision variables dependent on the time instant, the train and the infrastructure (Caprara et al., 2002). The objective function takes the cancellation and the deviation of the preferred timings into account.

In contrast to macroscopic timetabling, planning at the microscopic level guarantees a conflict-free timetable on the signaling level in ideal circumstances. A microscopic timetable thus provides a certificate of feasibility. This implies that the timetable can be plainly implemented in practice, but it is more complex to construct a microscopic timetable since many more constraints have to be taken into account. While the PESP is suitable to model the microscopic timetabling problem, no results hereof are published. Preliminary experiments on the microscopic level in the context of Chapter 4, have shown that it is more efficient to use auxiliary reservation and release times to construct a microscopic timetable than to use the PESP approach. The full explanation of this model can be found in Chapter 4.

On the one hand, existing literature on microscopic timetabling focuses on how to check a macroscopic timetable for feasibility at the microscopic level and how to (optimally) adapt this macroscopic timetable to obtain this feasibility in case it is not microscopically conflict-free (e.g. Bešinović et al., 2016a, 2017, Sels et al., 2016a, Caimi et al., 2011). These approaches usually construct the corresponding routing plan at the same time as the microscopic timetable. To avoid repetition and overlap, these approaches will be discussed in the section on integration of railway routing and timetabling.

On the other hand, there also exist approaches that start from an existing microscopic timetable and strive to improve the robustness of this initial timetable. Fischetti et al. (2009) for example, uses a set of delay scenarios in order to improve the robustness. Other approaches focus on estimating, determining or improving the amount or placement of supplements in a timetable. They achieve this by making only (relatively) small changes to the initial timetable while sticking to the initial routing plan. Andersson et al. (2015) and Kroon et al. (2008a) improve the location of buffer times and/or the distribution of supplements in the timetable at a specific point in the network or for a single train on a number of consecutive trips, respectively. As a result of the many interactions between trains in complex and dense station areas these approaches are not straightforwardly applicable to the networks considered in this dissertation. Optimization of running time supplements and their allocation is also studied by for example Rudolph (2003) and Yuan and Hansen (2007). Caimi (2009) divides the network in *condensation zones*, where capacity is limited, e.g. dense stations, and *compensation zones*, where the traffic is less dense, e.g. corridors. His strategy is to remove time reserves, i.e. supplements, to the compensation zones. Since train routes of many trains can cross or overlap several times in large, complex and dense railway stations close to each other, as are considered in this dissertation, there is a high risk for delay propagation. This necessitates the inclusion of time reserves within these areas, such that the strategy of Caimi (2009) does not suffice.

The advantage of approaches that only make small changes to an existing timetable is that they can have a relatively large impact on the performance of the railway system and that these changes can easily be implemented in practice without affecting the passengers too much. The disadvantage is that the final result is still highly dependent on the initial timetable. The approach proposed in Chapter 4 is devised to overcome this disadvantage. It strives for a globally optimal solution without any bias towards solutions that are similar to an initial input timetable. The emphasis of 'global' is on the fact that an exact optimization algorithm is used that looks for the global optimum. This is in contrast to a heuristic, that can be stuck in a local optimum and may output this local optimum as final solution. However, since the optimization will be restricted in computation time, the final solution is not assured to be the optimal solution. The emphasis of 'global' is not on the integrality of the network that is considered. In that sense, the scope is here restricted to the railway bottleneck and does not cover the whole railway network.

2.3.3 Line planning

Since line planning for rail takes the physical rail network as a fixed input, and provides a fixed input to subsequent timetabling, railway routing and rolling stock planning, assumptions can potentially be made about the form or characteristics of timetables, routing plans, rolling stock and rolling stock planning. Schöbel (2012) gives an overview of different approaches to model and solve the line planning problem, broadly categorizing line planning approaches that are (operator) cost-oriented, and those that are passenger-oriented.

Goossens et al. (2006) focus on minimizing operator cost, for the less-studied case of line planning where lines may not stop at every station. Also in Chapter 5, the stopping pattern of a line is decided upon in the line planning problem. The advantage of allowing lines to skip stations is the potential to combine fast lines which only stop at the stations with high demand and slow lines which also stop at stations with low demand (with the classification of stations not specified beforehand but decided upon during line planning). Using fast lines shortens the travel time of a lot of passengers and the slower lines assure that all stations will be served.

With a passenger focus, a common objective function is to maximize the number of direct travelers, i.e. the number of passengers who have a route from their origin to destination that does not require transfers. The simplest interpretation of this objective is to count the number of passengers for which there exists a line in the solution visiting both their origin and destination. This does not actually find passenger routes and does not guarantee that all counted passengers can actually *use* the line, as there may be insufficient capacity on some lines. In some networks, using this objective also has the risk that the passengers with no direct route may face many transfers. Another disadvantage is that maximizing the number of direct travelers encourages long train lines and does not favor skipped stations. This latter is critical in our case, since skipping stations is useful for making fast lines. Bussieck et al. (1997) is one example which uses this direct traveler objective, while ensuring that direct lines also have sufficient capacity to accommodate the passengers.

Another objective function with passenger focus is a travel time objective that takes into account the passenger's time traveling in trains combined with a penalty for switching trains (transfers). The calculation of this objective requires knowledge on the routing of passengers in the network taking into account travel time and transfers. This routing of the passengers can be modeled as paths in a graph, potentially requiring one path for every pair of stations. Schöbel and Scholl (2006) and Borndörfer et al. (2007) are examples where passengers are routed between a pair of stations by minimizing the sum of the travel time costs of the used paths. This passenger routing objective could be used as part of a weighted sum objective along with some operator cost (Borndörfer et al., 2007), or used alone but with an additional operator cost budget constraint (Schöbel and Scholl, 2006). In some practical problems, the inclusion of a budget constraint can be very important when combined with a passenger-oriented objective, as without it, solutions can contain many lines that individually satisfy every type of passenger. The line planning model proposed in Chapter 5 also uses the passenger's travel time objective. There are, however, tight rate limits on the maximum number of trains turning at a terminal station and on the use of certain infrastructure considered. Thus even without an operator budget consideration, there is no risk that solutions will contain particularly many lines.

Operator focused versus passenger focused is a first partitioning that can be made. Another partitioning is that a line planning model may be based on a predetermined set of lines (a line pool) or that it may find new lines dynamically. An advantage of a predetermined line pool is that all lines in the pool are guaranteed to be feasible in terms of line planning requirements. In Chapter 5 it is even required that the lines in the line pool are feasible in terms of timetabling requirements. A predetermined pool also has the advantage of limiting the problem size in a useful and dynamic way (because the pool can be limited to be as diverse or as focused as desired). However, it has the disadvantage that the full set of possible lines may be so large that enumerating them all would be intractable, while taking only a subset of all possible lines risks missing out on good solutions. Schöbel and Scholl (2006) present a model that takes as input a predetermined pool of lines. By contrast, Borndörfer et al. (2007) present a method where lines are generated dynamically as a pricing problem. This pricing problem finds maximum-weighted paths in an infrastructure network to be included as lines in a restricted master problem. However, the master problem itself is formulated in terms of a known line pool.

With respect to decision variables, many approaches are similar in using either a binary decision for the presences of each line, or a non-negative or integral decision for the frequency of each line, where a frequency of zero means that the line is not in the solution. In the approach in Chapter 5 only one of a set of frequencies, defined individually for each line, may be selected, so the proposed model uses a binary decision variable indicating the presence of a (line, frequency)-pair.

Related specifically to the problem that is addressed at DSB S-tog in Chapter 5, Rezanova (2015) solves the line planning problem with an operator focus, considering train driving time and a particular competing objective related to new regulations for drivers. Rezanova (2015) notes the problem of finding line plan solutions that are not feasible for timetabling, and suggests that an integrated approach would be valuable.

Overall, the modeling approach in Chapter 5 is similar to the work of Schöbel and Scholl (2006) in the construction of the graph for passenger flows, but differs in the capturing of frequency-dependent costs for passenger travel times. Also the line frequency is modeled in a stricter manner, which is necessary for the case study that is dealt with. In this case study, specific sets of frequencies are valid for each line where by contrast, in other work such as Schöbel and Scholl (2006) or Borndörfer et al. (2007) frequency is modeled as a discrete variable over all positive integers for each line.

2.3.4 Integrated methods

Line planning and timetabling

The research proposed in Chapter 5 is not the first attempt towards an integration of line planning and timetabling in railway scheduling. In Liebchen and Möhring (2007), some line planning decisions are included in the timetabling process. They assume that, for some parts (sequence of tracks) of the network, the number of lines serving each part is known beforehand. On these track sections they put an artificial station in the middle. Every line along this track section is then partitioned into two line segments, before and behind the artificial station. They use a PESP to model the timetabling problem to which they add constraints such that a perfect matching between the arriving and the departing line segments is forced. This is achieved by matching arrival and departure times of the line segments in the artificial station which are assigned by this same model. Here one line corresponds to one train. This approach is deficient if, for some network parts, the number of passing trains is not known beforehand. This is often the case in real world networks.

Kaspi and Raviv (2013) present a genetic algorithm that builds a line plan and timetable from scratch. They start from a given line pool and, per line, from a fixed number of potential trains. A solution consists of three characteristics for each train: the value zero or one, indicating whether the train should be scheduled or not, an earliest start time and a stopping pattern. A member of the initial population is constructed by drawing values for each characteristic from separate Bernoulli distributions. The timetable and line plan are constructed by scheduling trains with value one for the first characteristic according to a fixed priority rule. If a train cannot be scheduled without one or more conflicts with other already scheduled trains, this train is discarded from the solution. For the resulting timetable, the passenger travel time and the operator cost are calculated. These performance results affect the distribution parameters of the Bernoulli distributions from which the next generation will be drawn. This approach uses the performance of the timetable as input for the line planning of the next iteration. This interaction between line planning and timetabling is also the case in the approach proposed in Chapter 5. But in contrast to the stochastic approach of Kaspi and Raviv (2013), in Chapter 5 information of the timetable is used to make some deterministic and tactical changes to the line plan. Also in Goerigk et al. (2013) timetable performance is used to evaluate line plans. However, they do not iterate between the construction phase of the line plan and the timetable and they do not use this information to improve the line plan. They only use it to compare different ways to construct a line plan.

Michaelis and Schöbel (2009) offer an integrated approach in which they reorder the traditional sequence of line planning, timetabling and vehicle scheduling for bus planning. The different planning steps are, however, performed one after one such that the approach is still sequential. Vehicle scheduling or rolling stock scheduling are not integrated in this dissertation, but in Chapter 5 turn restrictions in the terminal stations are taken into account which significantly simplify the rolling stock scheduling. Taking turn restrictions into account is useful if terminal stations are not equipped with enough shunting space for efficient turning during daily operation. In fact, neglecting turn restrictions can lead to infeasible timetables. A thorough search of the relevant literature yielded that no other integrated approach for timetabling and line planning takes turn restrictions during daily operation into account.

Recently, Schöbel (2015) published a mixed integer linear program (MILP) in which line planning and timetabling are integrated for railway planning. This model is based on the PESP of Serafini and Ukovich (1989). In the model, binary variables are introduced to indicate whether a certain line is added to the line plan. There are also big M-constraints added to the PESP model in which these binary variables are used to push event times of lines which are not in the line plan to zero and also to switch off lower bounds of activities for unassigned lines. The objective function minimizes the planned travel time of the passengers. Transfer penalties are not taken into account, but they can easily be introduced as a weight in the objective function. No performance results of this model are presented yet.

Chapter 5 proposes a heuristic approach in which a line planning and timetabling module alternate, where each consists of an exact optimization model. An added value of this approach is that passenger robustness is taken into account when constructing a line plan (and timetable). The approach shifts the focus in current research from the integration of line planning and timetabling to the creation of passenger robust line plans (and timetables). This algorithm constructs a line plan that minimizes planned passenger travel time and operator costs but also prevents unreliable travel times during daily operation in order to provide a short travel time in practice for all passengers. As mentioned in Section 2.2, a *passenger robust* plan minimizes this total travel time in practice. In order to obtain short travel times in practice, the propagation of delays from one train to another has to be avoided, among other things. This can be achieved if the line plan allows a timetable with well-placed and large enough buffer times between trains. Also in Kroon et al. (2008b), Caimi et al. (2012), Salido et al. (2012), Dewilde et al. (2013), Sels et al. (2016a) and Vansteenwegen et al. (2016) the (minimum) buffer times between train pairs are lengthened in order to reduce the propagation of delays.

Another added value of the approach proposed in Chapter 5 is that trains with the same line are equally spread over the period of the cyclic timetable. Making the reasonable assumption that passengers arrive uniformly in a station of a high frequency network, a fixed time interval between any two subsequent trains of the same line, minimizes the average waiting time of the passengers before boarding.

Railway routing and timetabling

As is the case for timetabling approaches, there exist integrated approaches that start from an initial schedule and other approaches that start from the ground up. A minority of the integrated approaches start from an initial schedule. Examples that start from an initial schedule with a large positive impact on the performance of the railway system are Carey and Lockwood (1995), Dewilde et al. (2013, 2014a) and the approach presented in Chapter 3. These latter research achievements start from an initial routing plan and timetable and iterate between a routing module, a timetabling module and a platforming module in order to improve the passenger robustness of the schedule. The added value of the approach presented in Chapter 3 is that it takes passenger numbers and recurring delays explicitly into account in timetabling and railway routing without using a stochastic approach as in Kroon et al. (2007b), Meester and Muns (2007) and Kroon et al. (2008a). These latter approaches use a simulation module in the optimization algorithm, which makes the performance highly dependent on the available computation time. In Chapter 4 the same objective as in Dewilde et al. (2013, 2014a) and Chapter 3 is premised, but the approach starts from the ground up to obtain better results. This method consists of a separate routing model and timetabling model without a feedback loop between the models. So it is not immediately obvious that this method is categorized as an integrated method, but rather as a sequential method. The routing model constructs a routing plan with the design of a timetable by the timetabling model in mind: the intensity with which a node is used determines the maximum possible buffer time in that node and thus the vulnerability of that node for propagating delays. Spreading out the node usage as much as possible enables and facilitates the construction of a passenger robust timetable in the next step.

The existing approaches that solve the integrated (or sequential) problem usually combine a railway routing and a timetabling approach and optionally a feedback loop between the routing and the timetabling phase (Caimi et al., 2011, Schlechte et al., 2011, Goverde et al., 2016, Bešinović et al., 2016a, 2017, Sels et al., 2016a, e.g.). Most methods iterate therefore between the microscopic and the macroscopic level (Schlechte et al., 2011, Goverde et al., 2016, Bešinović et al., 2016a). In Chapter 4, by contrast, all scheduling is performed at the microscopic level. The method considers first the construction of the routing plan and thereafter a microscopic timetable is constructed. The order in which the routing plan and the timetable are constructed differs from the existing approaches and has the advantage that the trains can be spread over the available infrastructure without being restricted by a timetable. This approach proves worthwhile in dense networks with a complex infrastructure lay-out. This is the case because for complex railway station areas with many switches and route options, constructing a timetable on the signaling level seems the only way to avoid many trial and error iterations and to reach an optimal infrastructure usage of these areas.

Bešinović et al. (2016a), Goverde et al. (2016) and Sels et al. (2016a) constructed integrated approaches for railway routing and timetabling by using an interaction between the macroscopic and the microscopic level.

In (Goverde et al., 2016, Bešinović et al., 2017), an advanced integration of a microscopic and a macroscopic model is developed in order to construct a feasible, stable and robust timetable, where the macroscopic parameters are iteratively updated by recomputing them at the microscopic level. The authors detect places that are sensitive for delay propagation, for example based on an infrastructure occupation rate higher than the recommended threshold by (UIC, 2013). They manage delay propagation by including or removing supplements to dwell and running times of the appropriate trains, but this can also cause new conflicts. These detected conflicts are excluded by going back to the macroscopic level and assigning new arrival and departure times while dealing with this information. The final timetable is conflict-free at the microscopic level. The main difference with the research presented in Chapter 4 is the infrastructure lay-out and train density for which the method is designed. The approach of Bešinović et al. (2016a, 2017) works well and fast for a network with 40 trains and 26 stations. However, in their considered network at most four tracks lie in parallel and the maximum occupation rate is 54.7%. By contrast, Brussels-South and Brussels-North both have over 10 tracks in parallel and there pass 85 trains through Brussels during peak hour. The percentage of time that the platforms are blocked in Brussels station area is over 50% with a maximum of 77.3%. In fact, by restricting our network to the platform areas, we calculated a lower bound on the maximum occupation rate of Brussels station area, since we did not take the interactions and differences in speed limits in the grid zones into account for compressing the timetable. Another difference is that Bešinović et al. (2016a, 2017) and Goverde et al. (2016) use heuristic approaches, while the approach in Chapter 4 makes use of exact optimization models. The advantage of an exact approach is that the routing plan and the timetable can be optimally designed for the bottleneck. However, due to the complex infrastructure lay-out and the train density, the computation time for calculating an exact solution is often prohibitively large.

Sels et al. (2016a) developed an approach to plan a timetable at the macroscopic level and thereafter assign routes to the trains at the microscopic level. During the timetabling phase the impact of daily delays on the performance of the timetable is taken into account by using the passenger travel time in practice as the objective function. The passenger travel time in practice is estimated by a function that depends on the minimum possible travel time for each passenger and the added supplement to each passenger's trajectory. The shape of this estimation function depends on the expected primary delay functions of the (involved) trains. The right amount of supplements to add is decided upon by the model based on the estimate of the passenger travel time in practice, i.e. the corresponding function value of the estimation function. However, this approach does not assure the existence of a conflict-free assignment of routes. Complex station areas constitute a bottleneck for this approach. Any conflicts have to be solved by trial and error with small adjustments to the timetable. Moreover, an efficient infrastructure usage is not taken into account in this approach. This work is constructed for and validated on the Belgian railway network and optimizes passenger robustness. The research described in Chapters 3 and 4 is also validated on a part of the Belgian railway network. These chapters focus on the bottleneck, Brussels, while Sels et al. (2016a) and Sels (2016b) focus on a larger part of the network. Sels et al. (2016a) and Sels (2016b) get overall good results but less so when specifically looking at the bottleneck. Another difference is that these authors explicitly constrain the buffer times between different trains on a common node by imposing headway times, while the buffer times are not constrained in the approach in Chapter 3 and 4, but maximized in the objective function of the optimization model. This is possible because of the microscopic level that is worked at.

Lamorgese et al. (2016) propose an exact model to solve the railway routing and the timetabling problem at the same time, but they use a micro-macro heuristic to actually calculate the routing plan and the timetable. The number of route alternatives in their case study is small and they do not take any form of robustness into account. By contrast Caimi et al. (2011) propose a multi-level framework for generating train schedules in highly utilized networks. Their approach works fast, but they encounter the problem that the level of detail used to create the macroscopic schedule highly affects the result. With the problems they encounter, it seems hard to find an appropriate macroscopic network lay-out that simplifies the complex microscopic infrastructure lay-out considered in Chapter 3 and 4 of this dissertation. Furthermore, robustness is not considered either in their work.

Other integrated or sequential approaches do not use a feedback from the microscopic to the macroscopic level. In these approaches arrival and departure times are assigned on the macroscopic level and thereafter a routing plan is developed. The microscopic timetable which is created thereby is tested on its conflict-freeness, e.g. by simulation or an analytic approach. Research belonging to this category is for example Zwaneveld et al. (1996, 2001), Kroon et al. (2009), Schlechte et al. (2011) and De Fabris et al. (2013).

2.4 Conclusion

This chapter introduces railway terminology on the building blocks of a railway network and a railway planning. It explains how a train drives through a railway network and it introduces useful visualizations to represent a network or a schedule. A framework for the different planning levels is outlined. Railway planning is a complex task, that is usually divided into several subproblems. This dissertation presents developments for line planning, timetabling and railway routing. Several objectives can be premised when making the planning. From a passenger point of view, short travel times in practice, even in case of frequently occurring small delays, are preferred. This is captured in passenger robustness, which is the guiding principle in this dissertation. Different interpretations of robustness and how they can be implemented in practice are discussed in this chapter. Furthermore, this chapter situates the research presented in this dissertation in the state of the art on line planning, timetabling, railway routing and integrated approaches. The research here distinguishes itself by premising passenger robustness, by taking passenger numbers and recurring delays explicitly into account and by first focusing on the spreading in space to pave the way for a good spreading in time. The main idea in this dissertation is to come up with solutions in this research area that provide the best possible service to the passengers.

3

Improving passenger robustness on the tactical level

This chapter considers nearly saturated station areas whose limited capacity is one of the main reasons of delay propagation.[1] By taking action during the planning phase, the goal is to improve the total travel time in practice of all passengers in the railway network. Therefore, passenger numbers and recurring delays of trains are explicitly taken into account in railway routing and timetabling. By maximizing the (passenger- and/or delay-) weighted spreading between trains, potential conflicts that affect many passengers are avoided. This is beneficial for the total real travel time of all passengers. Using the presented approach, the passenger robustness of Brussels, Belgium's main railway bottleneck, can be improved by almost 4% compared to the existing literature (Dewilde et al., 2013, 2014a, Dewilde, 2014b).

[1]This chapter largely corresponds to the peer-reviewed conference paper of the RailTokyo2015 conference:

Sofie Burggraeve, Thijs Dewilde, Peter Sels, and Pieter Vansteenwegen. Improving passenger robustness by taking passenger numbers and recurring delays explicitly into account on the tactical level. In *Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015), Tokyo, Japan*, 2015b.

Many sentences are cited from this paper, but in order to keep the text easily readable, cited sentences are not separately indicated.

3.1 Introduction

The goal is to improve the passenger robustness of an existing timetable and routing plan. As mentioned in the literature review, the approach of Dewilde et al. (2013, 2014a) and Dewilde (2014b) serves this same goal. They developed an algorithm that indirectly improves the passenger robustness of an existing timetable and routing plan by making small changes to an existing timetable and routing plan. This results in only a small cost for the railway operator and the railway infrastructure manager. This chapter extends their approach and presents a large positive impact on the passenger robustness. Since (Dewilde, 2014b) is a PhD thesis encompassing the information of Dewilde et al. (2013) and Dewilde et al. (2014a), the former will be the only text that will be referred to from now on, in order to avoid excessive referencing. The main contributions of this chapter are:

- A way to use passenger data while improving the passenger robustness of an existing timetable.
- A way to use historical delay data while improving the passenger robustness of an existing timetable.
- A validation of both approaches separately and of combinations of both approaches on a realistic, large and complex railway bottleneck, namely Brussels (Belgium).
- Resulting routing plans and timetables for Brussels, which are more passenger robust than reference routing plans and timetables from practice and from the literature.

The approach of Dewilde (2014b) is described first, in Section 3.2. Subsequently, the modifications to this approach are explained in Section 3.3 and 3.4. In Section 3.5, the simulation tool is introduced with which the performance of a timetable and routing plan is determined in this dissertation. Section 3.6 specifies the case study on which the approach is validated. Thereafter, the results are presented in Section 3.7. Section 3.8 concludes this chapter and Section 3.9 contains ideas for further research.

3.2 Approach of Dewilde

This section describes the approach of Dewilde (2014b). The algorithm of Dewilde (2014b) needs an existing railway timetable and routing plan as input.

It consists of three modules: the routing module, the timetabling module and the platforming module. These modules iteratively improve the spreading in the routing plan and the timetable, in space and time respectively. Afterwards, the resulting timetable and routing plan are evaluated with a simulation tool. Figure 3.1 shows an overview of the algorithm of Dewilde (2014b).

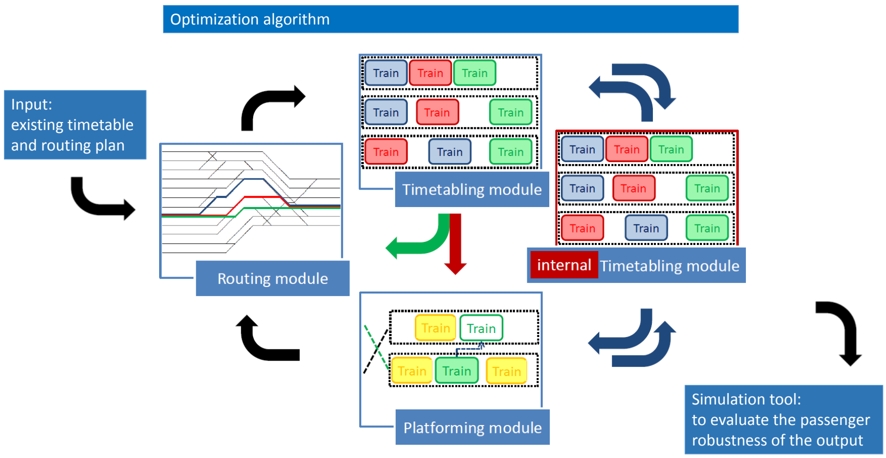


Figure 3.1: Overview of the optimization algorithm of Dewilde (2014b). This figure is based on Dewilde (2014b).

3.2.1 Objective function

Instead of directly optimizing passenger robustness, Dewilde (2014b) strives towards an increase in the shortest buffer time between each pair of trains. He argues that directly optimizing passenger robustness would make the approach complicated and time consuming, since this would require knowledge on real travel times of passengers during optimization, for which delay propagation computations would be needed. The shortest buffer time between two trains is measured at the instant when the two trains drive the closest to one another when driving in the same direction. When driving in opposite directions, the shortest buffer time is measured at the instant when the free time between the blocking times of both trains on the shared infrastructure, is the smallest. The shortest buffer time between two trains is also referred to as the minimum buffer time between two trains. Let T be the set of all trains considered in the railway system and assume that for every train a route is fixed. Let $B_{t,r,t',r'}$ be the shortest buffer time between train t with route r and train t' with route r'. The objective function can be represented by:

$$\min \sum_{\{t,t'\}\subset T} C_{t,r,t',r'}, \tag{3.1}$$

where t and t' represent two different trains ($t, t' \in T$), r is the route fixed for train t and r' is the route fixed for train t' and $C_{t,r,t',r'}$ is defined as

$$C_{t,r,t',r'} = \begin{cases} 15 & \text{if } B_{t,r,t',r'} \leq 0 \quad \text{(conflict);} \\ \frac{1}{B_{t,r,t',r'}} & \text{if } 0 < B_{t,r,t',r'} \leq 15; \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

If the routes r and r' have no common infrastructure, then $B_{t,r,t',r'}$ is assigned the value infinity. If the routing plan and the timetable cause a conflict between trains t and t', then $B_{t,r,t',r'}$ is assigned a negative value and $C_{t,r,t',r'}$ is set to 15. If it is clear to which routes r and r' train t and t' are assigned to respectively, then the notation $B_{t,r,t',r'}$ and $C_{t,r,t',r'}$ can be simplified to $B_{t,t'}$ and $C_{t,t'}$ respectively. The objective function minimizes in fact the sum of the inverse shortest buffer times. The inverse shortest buffer time is the cost associated with a train pair and a fixed set of routes for this train pair. Compared to maximizing the sum of the shortest buffer times, minimizing these costs has the advantage that increasing the smallest shortest buffer times with a certain amount of time improves the objective value more than increasing larger shortest buffer times with the same amount of time. This objective incorporates the idea that the smaller the buffer time, the larger the probability on propagation of delays, but also that the probability on propagation of delays quickly diminishes with the length of the buffer time. However, minimizing the sum of the inverse shortest buffer times, makes the objective function non-linear, which has computational costs.

To improve the value of this objective function, trains can be spread in space and time. For example, choosing another route for a train, can enable two trains to no longer overlap in space and thus that their minimal buffer time becomes infinite and the corresponding cost becomes zero. Another example is a small shift in arrival and departure times of a train, such that the minimal buffer time between this train and another train increases with the length of this shift. Of course, in order to really improve the objective function, one has to check whether the new route or the shift in arrival and departure times will not have a larger negative effect on the shortest buffer times between the adapted train(s) and other trains.

Note that in formula (3.2) the cost depends on the length of the buffer time, but is independent of the number of passengers on these trains and the typical delays of both trains.

3.2.2 Routing module

This module assigns the trains to routes such that the total cost in equation (3.1) is minimized. The input is a set of possible routes for each train and the arrival and departure times for each train in each station. Before the assignment, the shortest buffer time between each pair of possible routes of different trains must be calculated. The shortest buffer time is determined at the microscopic level. The assignment itself is done by an exact optimization model. In this model, conflict-freeness is imposed as a hard constraint. This model is now discussed in more detail.

Let R_t be the set of possible routes for train t. The variables $x_{t,r}$ are binary variables that indicate whether train t is assigned to route r or not. For each train t and route $r \in R_t$, the set $R^c_{t,r,t'}$ is defined as $\{r' \in R_{t'} : B_{t,r,t',r'} \leq 0\}$ and contains all routes of train t' that are conflicting with train t and route r.

$$\min \sum_{\{t,t'\} \subset T, r \in R_t, r' \in R_{t'}} C_{t,r,t',r'} x_{t,r} x_{t',r'}, \tag{3.3}$$

subject to

$$\sum_{r \in R_t} x_{t,r} = 1 \qquad \forall t \in T, \tag{3.4}$$

$$x_{t,r} + \sum_{r' \in R^c_{t,r,t'}} x_{t',r'} \leq 1 \qquad \forall t \in T, r \in R_t, t' \in T \setminus \{t\}, \tag{3.5}$$

$$x_{t,r} \in \{0,1\} \qquad \forall t \in T, \forall r \in R_t. \tag{3.6}$$

The cost function (3.3) is the objective function. It sums the costs of the assigned train pairs. Constraints (3.4) assure that exactly one route is assigned to each train. Constraints (3.5) prohibit that conflicting routes can be assigned. Constraints (3.6) are the type settings for the decision variables.

There are (at least) two graph-theoretic interpretations of the routing model. First, model (3.3) - (3.6) can be categorized as a node packing model, where the node packing constraints are represented by constraints (3.5). This can be explained as follows. For each combination of a train and a possible route for this train, a node is created in the graph. If two of such combinations cause a conflict, so $r' \in R^c_{t,r,t'}$, an edge that links the node associated with train t and route r with the node associated with train t' and route r', is included in the graph. In general, a *feasible solution to a node packing problem* is a node set for which no two nodes are linked with an edge. In the case of the routing model, we are looking for a solution that contains a node for every train, since all the trains have to be planned. Note that a solution can never contain two nodes associated with a single train, since two different routes for this train will always be conflicting. Their corresponding nodes will be linked with an edge. As a consequence they cannot both take part in a solution to the node packing problem. Furthermore, the optimal solution is a feasible solution that contains a node for every train and has the best value for the cost function (3.1). Secondly, model (3.3) - (3.6) can be solved by looking for k-cliques in k-partite graphs. Again, for each combination of a train and a possible route for this train, a node is created in a graph. The graph is partitioned by grouping the nodes according to the trains that they are associated with. So k equals the number of trains that have to be planned. Opposite as in the graph for the node packing representation, now two nodes are linked if they do not cause a conflict. So, there will never be a link between two nodes associated with a single train. A feasible solution is now a k-clique in this k-partite graph. By construction of the graph, such a k-clique will only contain combinations of trains and routes which assure a conflict-free timetable. The solution to the routing problem is the k-clique with the best value for the cost function (3.1). More information on this problem and how it can be solved can be found in (Grünert et al., 2002).

Unfortunately, the objective function (3.3) is non-linear. Therefore, we define for all train-route combinations (t, r) with $t \in T$ and $r \in R_t$ the non-negative, continuous variable $z_{t,r}$. These variables can be interpreted as the spreading cost associated with one train-route combination. Note that the $C_{t,r,t',r'}$'s are spreading costs associated with two train-route combinations. The $z_{t,r}$ variables are introduced to linearize the cost (3.3) according to the technique of Kaufman (Kaufman, 1978). Constraints (3.7) set a lower bound for these spreading costs:

$$\sum_{t' \in T, t' \neq t} \sum_{r' \in R_{t'}} C_{t,r,t',r'}(x_{t,r} + x_{t',r'} - 1) \leq z_{t,r} \qquad \forall t \in T, \forall r \in R_t. \tag{3.7}$$

So, if route r and route r' are selected for trains t and t' respectively, then the cost corresponding to their shortest buffer time contributes to the cost variables $z_{t,r}$ and $z_{t',r'}$. The alternative objective function then becomes:

$$\min \qquad \sum_{t \in T} \sum_{r \in R_t} z_{t,r}, \tag{3.8}$$

where

$$z_{t,r} \geq 0 \qquad \forall t \in T, \forall r \in R_t. \tag{3.9}$$

Routing plans with a better spreading between the trains lead to a better value of this cost function. Note that the cost of assigning route r to train t equals the sum of the costs associated with the shortest buffer time between this train and any other train. As a result, the optimal value of objective (3.8) will be twice the optimal value of objective (3.3).

3.2.3 Timetabling module

This module changes the arrival and departure times of the trains in order to improve the cost function in (3.1). As input for this module, train routes, assigned by the routing module, are used as are the arrival and departure times of the input timetable or the timetable of the previous iteration. Cleverly changing the arrival and departure times can increase the minimal buffer times between train pairs. In this module, also the order of trains may be changed in order to increase the minimal buffer times. The new arrival and departure times assigned by this module are constrained by a time window. This is to avoid that the time window in which the trains are planned increases when the minimum buffer times increase. Only a small shift in arrival and departure times is considered in each iteration. Dewilde (2014b) argues that taking all possible variations into consideration would make the problem (too) large for complex and dense railway station areas. That is why he uses a heuristic approach. He evaluates step by step the effect of (i) a shift in the arrival and departure times of one train, (ii) a combined shift of arrival and departure times of more trains and (iii) an order swap of two trains. Each of these three options to change arrival and departure times is referred to as a *move*. For each of these moves, it is possible that changing the arrival and departure times of one train causes a conflict with another train. Therefore, an *internal timetabling module* is present to check for a solution that resolves the conflict(s) before evaluating the entire move. This allows to check the potential improvement of a move. Such a solution can be found by considering multiple parallel changes of arrival and departure times. In case a move, possibly including conflict resolving, improves the cost function, the timetable is adapted. The resulting timetable does not contain conflicts. The timetabling module also contains a smallest ascent procedure in order to be able to escape from local optima.

3.2.4 Platforming module

This module may change the platforms of trains in order to improve the cost function (3.1). In the routing and the timetabling module, the platform of each train in each station is fixed. As input for this module, train routes, assigned in the routing module, are used as are the arrival and departure times of the trains in the stations. Based on the shortest buffer times, a set of trains is selected for which a platform change is considered in the platforming module. A platform change is carried out only if it improves the objective function (3.1). In order to assess the potential improvement of a platform change, also newly created timetabling opportunities are tested with the internal timetabling module. Changing the platform also changes the route towards and away from that platform.

Using a new platform and a new route towards and away from this platform improves the spreading in space. Since promising timetable changes are simultaneously taken into account, the spreading in time improves as well. Spreading in time and spreading in space both have a positive effect on the cost function (3.1). This makes the platforming module a valuable part of the algorithm (Dewilde et al., 2013).

3.2.5 Interaction between the modules

The algorithm starts by executing the routing module. Thereafter the timetabling module is started. If there is no improvement (or hardly an improvement) by changing the arrival and departure times and the orders of trains, then the platforming module is executed. Thereafter the routing module is started again. If the timetabling module produces a considerable improvement, then the platforming module is skipped. This loop (routing - timetabling (- platforming) - routing - $\cdots$) is repeated until there is no further improvement noticeable or until a certain number of iterations is reached. Figure 3.1 presents an overview of this algorithm. The internal timetabling module represents the timetabling module that is called inside the timetabling module or the platforming module to assess the potential improvement of timetable or platform changes.

It is straightforward to see that the number of passengers on a train and the typical recurring delays of trains influence the total passenger travel time in practice. However, this information on passenger numbers and recurring delays is not taken into account in the approach of Dewilde (2014b). In the following two sections, it will be explained how this information can be included in the framework of Dewilde (2014b). Thereafter a case study will show that when this information is taken into account as it is described here, the passenger robustness can be improved by up to 4% more.

3.3 Introducing passenger numbers

In order to improve the total real travel time in practice of all passengers, it seems advantageous to come up with a solution in which trains with more passengers are made less liable to delays than hardly occupied trains. This observation is used to improve the algorithm outlined above. This section first explains how passenger numbers will be taken into account and thereafter how and where exactly this will have an influence on the framework of Dewilde (2014b).

One way to make a crowded train less susceptible to delays is to increase the buffer time between such a train and the trains that use the same infrastructure just before this train. This can be included in the algorithm of Dewilde (2014b) by prioritizing these train pairs. From now on, the first train refers to the train that uses the shared infrastructure first. The second train refers to the train that uses the shared infrastructure later.

A well-known method to prioritize is to use weights. The costs $C_{t,r,t',r'}$ in the objective function (3.1) and (3.3) indicate the cost of choosing the corresponding train-route combinations. In this dissertation, we go one step further by cleverly weighting these costs. The more passengers (on the second train) that will be affected by a delay of the first train, the higher the weight on the cost of this pair of train-route combinations. Furthermore, the higher the weighted cost the more important it is to spread the trains in the corresponding train pair well. The weights have to reflect the amount of passengers whose real travel times have a higher probability to exceed the planned travel times due to propagation of a delay of the first train. Therefore the weights depend on (i) the number of passengers that travel with the second train at the instant the train uses the shared infrastructure and (ii) the number of passengers that board on the second train in the stations of the bottleneck beyond the common infrastructure. If there are different places in the network where both trains use the same infrastructure, the instant when the shortest buffer time between the two trains occurs, is chosen to count the passengers in order to determine the weights. If two trains do not share infrastructure, the weight is set to zero, just like the cost incurred by the ‘infinite buffer time’ between these two trains.

3.3.1 Implementation

This section explains first how the new weights are determined and thereafter how they are implemented in the algorithm of Dewilde (2014b). The passenger numbers that are used to determine the weights in the optimization algorithm coincide with the passenger numbers that will be used in the simulation afterwards.

The passenger weight for train-route combinations (t, r) and (t', r') will be indicated with $P_{t,r,t',r'}$. This weight coincides with the number of passengers on the second train at the instant that the shortest buffer time between the two trains occurs together with the number of passengers that will board on the second train after this instant. The place in the network where the shortest buffer time between a train pair occurs can be determined by computing the time instants at which these trains enter and leave the shared infrastructure. This is done at the microscopic level. The shared infrastructure depends on the routes considered for the train pair. At the place where the minimum buffer time is measured, the number of passengers that are on the second train or will board this train in an upcoming station can be determined based on the input passenger numbers.

Changing routes or arrival and departure times can all have a large effect on these weights. If for at least one of both trains another route is considered, the shared infrastructure can be different as well. Consequently, also the number of passengers can be different, as passengers can enter and leave the second train in stations in between both places. An example is provided in Figure 3.2. The number of passengers on the second train of train-route combinations (t, r) and (t', r') will be indicated as $P_{t,r,t',r'}$ (or as $P_{t,t'}$ if the routes are known and fixed). The situation in Figure 3.2 is now described.

Suppose that train A and B drive from left to right through the station. Train A releases the infrastructure in front of the station three minutes before train B reserves the (common) infrastructure in front of the station. The buffer time between A and B before they reach the station is three minutes. Suppose also that train A has a dwell time of one minute on its platform and train B has a dwell time of two minutes on its platform. Assuming that the blocking time of A is the same for the infrastructure in front and behind the station, the buffer time increases to four minutes behind the station. Thus, in the left picture, the shortest buffer time between both trains is three minutes and is achieved in front of the station. The assigned weight will be the sum of the number of passengers that are on train B at this instant and the number of passengers that will board on train B in the subsequent stations of the bottleneck. In the right picture, the shortest buffer time between trains A and B is four minutes and the critical place is located behind the station. The assigned weight will be the number of passengers on the train at that instant and the number of passengers that will board on train B in the subsequent stations. Both weights differ by the number of passengers that leave train B in the station. Note that selecting another route for one of the two trains can cause that the routes of both trains are no longer overlapping. Calculating buffer times in this case is no longer relevant and the weight becomes zero.

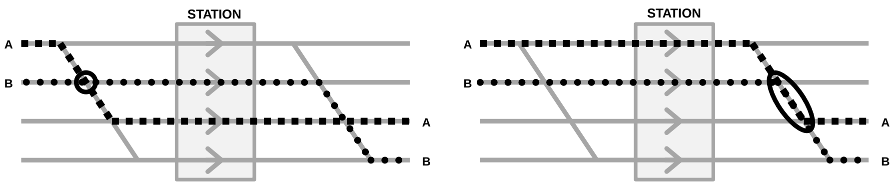


Figure 3.2: Example: Train A and B drive from the left to the right through the station. Suppose that the arrival time of A at the station is at time instant zero and that of B is at time instant three (in minutes), the dwell time of A is one minute and that of B is two minutes. Let r be the route of train A in front of the station, r' the route of train A behind the station and r'' the (entire) route of train B. The shortest buffer time in the left picture is reached in front of the station and in the right picture behind the station. Consequently, the weight $P_{A,r,B,r''}$ will differ from the weight $P_{A,r',B,r''}$ by the number of passengers that leave train B in the station.

Changing the arrival and departure times of the trains in the timetabling module can affect the weights, since the place where the shortest buffer time occurs can change. Changing the order of the trains always affects the weights, since the passengers on the new second train will become important.

The focus is now on the appearance of these weights in the algorithm. The definition of the cost in (3.1) must be adapted to include the passenger weights. Let $C^P_{t,r,t',r'}$ represent the passenger weighted cost associated with train-route combinations (t, r) and (t', r'). This passenger weighted cost is straightforwardly defined as

$$C^P_{t,r,t',r'} = P_{t,r,t',r'} C_{t,r,t',r'}. \tag{3.10}$$

By using this cost $C^P_{t,r,t',r'}$ instead of $C_{t,r,t',r'}$ in the cost functions (3.1) and (3.3), passenger weighted buffer times will be optimized. The higher $C^P_{t,r,t',r'}$, the less plausible that train-route combinations (t, r) and (t', r') will both be selected.

The (new) weights are not only used to measure the quality of new routing and timetabling solutions. They also play a role in the selection of the train pairs which are considered for timetable adaptations. In Dewilde (2014b), this selection is only based on the shortest buffer times. Here, this selection procedure is adapted by taking also the affected passengers on the second train into account. Train pairs considered for changes are selected according to their value of the new cost function (3.10). Furthermore, in the algorithm of Dewilde

(2014b), the selected train pairs are dealt with in a random order, while the train pairs are now ordered and handled based on their negative impact on the timetable. The selection of the trains considered for a platform change, is not altered.

3.4 Introducing recurring delays

Suppose that a single train regularly enters a restricted railway area with a delay. This section first explains why it is favorable to take this recurring delay explicitly into account in the planning of the restricted network. Thereafter is explained how these recurring delays can be taken into account in the matheuristic. Once again weights will be used and they will affect the framework of Dewilde (2014b) in the same way as described in the previous section.

Suppose first that the first train of a train pair is recurrently delayed. If the buffer times between this train and trains that use the same infrastructure shortly behind this train are shorter than the delay of the first train, the second train needs to be halted at the common infrastructure to avoid a conflict. The time the second train has to wait while the first train clears the common infrastructure, causes a lengthening of the real travel time for all passengers on that second train. Thus, it is important that these buffer times are large enough to avoid the propagation of this recurring delay. Here it is assumed that the amount of supplements is small compared to the recurring delays. So, buffer times that are not as large as the *mean recurring delay* of the first train are prioritized. Even in case the buffer time is equally long as the mean recurring delay, conflicts are not excluded in case the actual delay is larger than the mean recurring delay. Often, however, an exponential distribution is assumed for the delays (Goverde, 1998), for which it holds that the larger the mean recurring delay, the smaller the probability to draw a delay that is larger than the mean recurring delay.

Suppose now that the second train of a train pair is recurrently delayed when it enters the restricted network. The buffer time between the two trains is lengthened by the time length of the delay. So it becomes less urgent to deal with that buffer time in the optimization algorithm.

It can be concluded that not only the recurring delay of the first train has an impact on the buffer time between a train pair, but that also the recurring delay of the second train has to be taken into account. A recurring delay of the first train makes the buffer time between a train pair more critical, while a recurring delay of the second train makes it less critical (Sels et al., 2013). We want to include both opposing effects in the optimization algorithm. In much the same way as was the case when introducing passenger numbers, new weights will be used to indicate the importance/need of lengthening a certain buffer time. The associated cost will get a weight bigger than one if it becomes more important to lengthen the corresponding buffer time, e.g. if the first train is recurrently delayed but the second train not. The associated cost will get a weight smaller than one if it becomes less important to lengthen that corresponding buffer time, e.g. if the second train is recurrently delayed but the first train not. In each case, the magnitude of the weight will depend on the mean recurring delay of the trains. In case both trains of a train pair are recurrently delayed, the mean recurring delay of both trains determines the weight. For example, if the mean recurring delay of the first train is larger than that of the second train, this increases the probability of a conflict to occur. Thus the buffer time becomes more critical and the cost will get a weight bigger than one.

3.4.1 Implementation

Let D_t be the mean recurring delay for train t. We start with some examples that offer an interpretation of the new weights that will be included to take the effect of recurring delays into account. Suppose that the buffer time between two trains t and t' is five minutes for certain routes r and r'. Train t uses the common infrastructure first and thereafter train t'.

Example 1 Suppose that the mean recurring delay of train t is one minute. Train t' is never delayed. The 'altered' buffer time is then on average 20% shorter: $\frac{D_t}{B_{t,r,t',r'}} = \frac{1}{5} = 20\%$. This holds for all routes r and r' assigned to train t and t' that share infrastructure. We could now find it $\frac{1}{1-20\%}(> 1)$ times more important to increase the buffer time between t and t'.

Example 2 Suppose that train t' has a mean recurring delay of two minutes and train t is never delayed. Then the altered buffer time is 40% longer, such that we could find it $\frac{1}{1+40\%}(< 1)$ times less important to further increase this buffer time.

Example 3 Suppose now that train t has a mean recurring delay of one minute and train t' has a mean recurring delay of two minutes. Then the buffer time between t and t' is on average lengthened with 20%. We could find it now $\frac{1}{1+20\%}(< 1)$ less important to further increase this buffer time.

Let $D_{t,r,t',r'}$ be the relative importance representing how much more or less important it is to increase the planned buffer time between train-route combinations (t, r) and (t', r'). The relative importance derived in the previous three examples, can be generalized and embedded in one formula:

$$D_{t,r,t',r'} = \frac{1}{1 - \frac{D_t - D_{t'}}{B_{t,r,t',r'}}}, \tag{3.11}$$

where it is supposed that train t uses the common infrastructure before train t' does. Again the definition of the cost in (3.1) must be adapted to include these new weights. Let $C^D_{t,r,t',r'}$ represent the adapted cost that weighs recurring delays associated with train-route combinations (t, r) and (t', r'). This weighted cost is straightforwardly defined as

$$C^D_{t,r,t',r'} = D_{t,r,t',r'} C_{t,r,t',r'}. \tag{3.12}$$

By using this cost $C^D_{t,r,t',r'}$ instead of $C_{t,r,t',r'}$ in cost functions (3.1) and (3.3), weighted buffer times will be optimized, where the weights are determined by the recurring delays.

We add the restriction that the weighted cost should never be bigger than the cost of a conflicting train pair:

$$\forall t, t' \in T, r \in R_t, r' \in R_{t'} : C^D_{t,r,t',r'} \leq 15. \tag{3.13}$$

Formula (3.11) is not meaningful in the following cases:

- if $B_{t,r,t',r'} \leq 0$. Then $D_{t,r,t',r'}$ is assigned the value 1, such that the weighted cost $D_{t,r,t',r'} C_{t,r,t',r'} (= 1 \cdot 15 = 15)$ remains maximal;
- if $B_{t,r,t',r'} > 0$ and $\frac{D_t - D_{t'}}{B_{t,r,t',r'}} > 1 \Leftrightarrow B_{t,r,t',r'} - D_t + D_{t'} < 0$. Then the mean recurring delay of the first train exceeds the aggregated buffer time and the mean recurring delay of the second train. Both trains are in conflict when the mean recurring delays are taken into account. Thus assign to $D_{t,r,t',r'}$ the value $15 \cdot B_{t,r,t',r'}$, such that $D_{t,r,t',r'} C_{t,r,t',r'} = 15 \cdot B_{t,r,t',r'} \frac{1}{B_{t,r,t',r'}} = 15$, which equals the cost of a conflicting train pair;
- if t and t' with routes r and r' respectively have no common infrastructure. The value of $D_{t,r,t',r'}$ is of no importance as $C_{t,r,t',r'} = 0$ in this case, assign the value one to $D_{t,r,t',r'}$.

The interpretation of the fact that the planned buffer times are weighted with $D_{t,r,t',r'}$ in the matheuristic now becomes straightforward. The priority of buffer times that are at risk to be decreased by the impact of the recurring delays, will increase. The priority of buffer times that are lengthened by the impact of the recurring delays, will decrease. The newly updated objective function where the impact of the passengers is neglected, but the impact of the recurring delays is taken into account, becomes:

$$\min \sum_{\{t,t'\}\subset T, r\in R_t, r'\in R_{t'}} C^D_{t,r,t',r'}. \tag{3.14}$$

Note that if $D_{t,r,t',r'}$ and $C_{t,r,t',r'}$ are replaced with their formulas, (3.11) and (3.2) respectively, for a train pair with $0 < B_{t,t'}$ and $B_{t,t'} - D_t + D_{t'} > 1$, one obtains:

$$D_{t,r,t',r'}C_{t,r,t',r'} = \sum_{\{t,t'\}\subset T, r\in R_t, r'\in R_{t'}} \frac{1}{1 - \frac{D_t - D_{t'}}{B_{t,r,t',r'}}} \frac{1}{B_{t,r,t',r'}} \tag{3.15}$$

$$= \sum_{\{t,t'\}\subset T, r\in R_t, r'\in R_{t'}} \frac{1}{B_{t,r,t',r'} - D_t + D_{t'}}. \tag{3.16}$$

The denominator in (3.16) represents the buffer time that is to be expected in between both trains if the mean recurring delays are taken into account. Thus minimizing the objective function (3.14) boils down to minimizing the sum of the inverse 'expected' buffer times instead of the sum of the inverse 'planned' buffer times.

These weights play the same role in the optimization algorithm as the weights based on the passenger numbers. The details can be found in Section 3.3.1.

3.5 Simulation tool

In this dissertation, the simulation tool of Dewilde et al. (2014a) is used to measure the performance of a railway timetable and routing plan and to compare the performance of different schedules. The main performance criterion that it measures, is the passenger robustness of a timetable and routing plan. The simulation tool is a discrete event Monte Carlo simulation where events consists of trains blocking their next section and in which (small) initial train delays are drawn from a probability distribution in each simulation run. The simulation tool performs 10 000 simulation runs. The probability distribution that is used in both Dewilde et al. (2014a) and in this dissertation is the exponential distribution. The exponential distribution is known to be a good approximation of historical delays (Goverde, 1998). Initial delays are generated by a simulation draw and therefore train trips cannot pass off as planned. This results in knock-on delays and prolonged real travel times. In the simulation tool knock-on delays are registered for each event when a train cannot block its next section when requested. The knock-on delay is then equal to the time that this train has to wait before it can block its next section. The number of passengers that enter, leave or remain in a train in a certain station is assumed to be known in advance for all trains and all stations of the considered network and is assumed to be independent of the timetable. Based on this information the passenger robustness of a timetable and routing plan can be calculated in each simulation run. As defined earlier, the passenger robustness is a weighted sum of the travel times in practice over all passengers. The weights reflect how passengers perceive travel time (waiting in a station, driving in a train at full speed, etc.). The weights that are used to calculate the passenger robustness are listed in Table 3.1 (Dewilde, 2014b).

Table 3.1: Passenger robustness: weights (Dewilde, 2014b).

Kind	Weight
Nominal travel time	1
Out delay	3
Exit delay	3
Missed transfer	3
Used supplement	1
Non-used supplement	2

The out delay refers to the delay of passengers that leave a train inside the considered network. The exit delay refers to the delay of passengers that leave the considered network inside a train. The weight for a missed transfer is applied to the time that passengers have to wait for the next connecting train when they miss a planned transfer. A planned transfer is a transfer that is accounted for during timetable optimization. A used supplement refers to a supplement that a train used to absorb (part of) its delay. A non-used supplement refers to a supplement that appeared superfluous, meaning the train was not delayed.

The travel time in practice can easily be calculated for each simulation run, since each simulation run results in a value for the arrival time in the network and on the platforms, the departure time on the platforms and the exit time out of the network for all trains.

The simulation tool differentiates between the planned travel time, the passenger travel time in practice and the nominal travel time. The *planned travel time* is the sum of the weighted travel times of all passengers in ideal circumstances, implying all trains drive according to schedule. The passenger travel time in practice is the sum of the weighted travel times of all passengers as observed in practice. Unfortunately, circumstances in practice are often non-ideal, since the train system often has to deal with delays. These circumstances are imitated in the simulation by drawing delays from a delay scenario. The *nominal travel time* is the sum of the travel times of all passengers for a schedule in ideal circumstances (absence of delays), without weighing the travel times with the weights of Table 3.1.

3.6 Case study

The approach is tested on the case study of Brussels, Belgium's main railway bottleneck. First the network and the line plan used as input for the construction of the routing plan, are described. Then, the computation of the blocking times that are used to exclude conflicts, is explained. Thereafter, the generalizability of this case study is discussed. This includes the description of other sources that developed a timetable and routing plan for this same case study. This section is concluded with the delay scenario that is used in the simulation tool to measure and compare the performance of the different routing plans and timetables for this case study. A delay scenario describes for each train the delay distribution, including the distribution type, the amount of trains that are delayed and the determining parameters of the distribution (e.g. the average for the negative exponential distribution). In each simulation run, delays are then drawn from the distributions described in the delay scenario.

3.6.1 Network

The case study contains the grid zones and platform areas between Brussels-South and Brussels-North and it also includes the beginning of the open tracks, the outer grids and the entrances to the shunt yards. This area will be referred to as Brussels station area. It contains three out of five of Belgium's busiest stations and is a real bottleneck[2]. The 19 platform tracks of Brussels-South are connected via three tunnels with six platform tracks in Brussels-Central to twelve platform tracks in Brussels-North. Moreover, Brussels station area also includes the stations Brussels-Kapel, Brussels-Congres and Brussels-Schaarbeek with respectively six, six and twelve platforms. In addition to the tracks to the shunt yards, Brussels-South and Brussels-North have several tracks that lead trains from all over the country to and from Brussels, which necessitates a lot

[2]Source: https://pvmagazine.nl/brussel-noord-nu-het-drukste-station-van-belgie/, consulted in July 2016

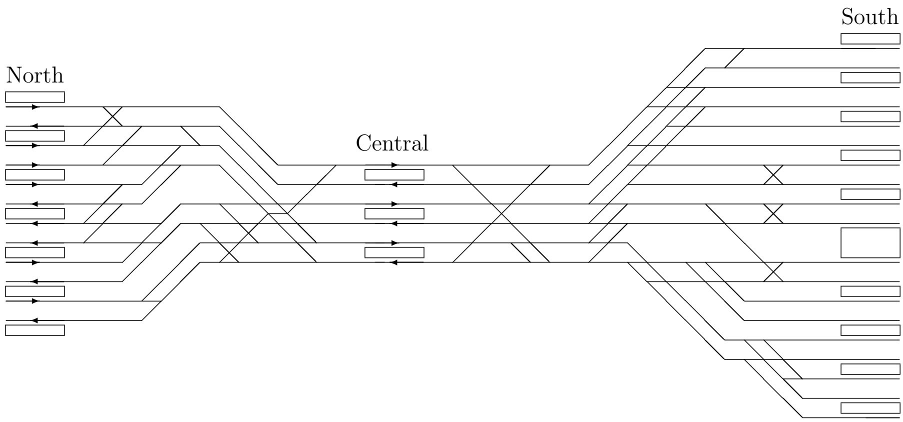


Figure 3.3: The core of Brussels dense railway area (Dewilde, 2014b).

of switches in this station area. Considering every border point, switch and platform as a node in the network, this leads to 481 nodes. There are 419 signals in this network. The core of the considered area is represented in Figure 3.3. A schematic overview of the entire station area is presented in Figure 3.4[3].

3.6.2 Line plan

The line plan, used as input for the approach, concerns 85 trains that pass through Brussels during peak hour in the morning (7 a.m. - 8 a.m.). These trains almost saturate the network, which makes this area interesting to test the approaches presented in this dissertation. Trains can enter Brussels-South from six different incoming lines and they can leave Brussels-South via seven different outgoing lines. Trains can enter Brussels-North from seven different incoming lines and they can leave Brussels-North via eight different outgoing lines. There are 73 trains traversing the whole network (from border node to border node). The other 12 trains start or end their route in a platform area. Six of these 12 trains enter or leave Brussels station area without passing Brussels-North, Brussels-Central and Brussels-Schaarbeek. So their trip ends or starts an outward line in Brussels-South. There are 42 trains driving in the direction from South to North and the other 43 trains are driving in the direction from North to South. There is one train that splits into two trains in the platform area of Brussels-South. There are also three trains that end

[3]Source: http://www.infrabel.be/sites/default/files/documents/ns_c-01-map-net-1045901_1.pdf, consulted in September 2014.

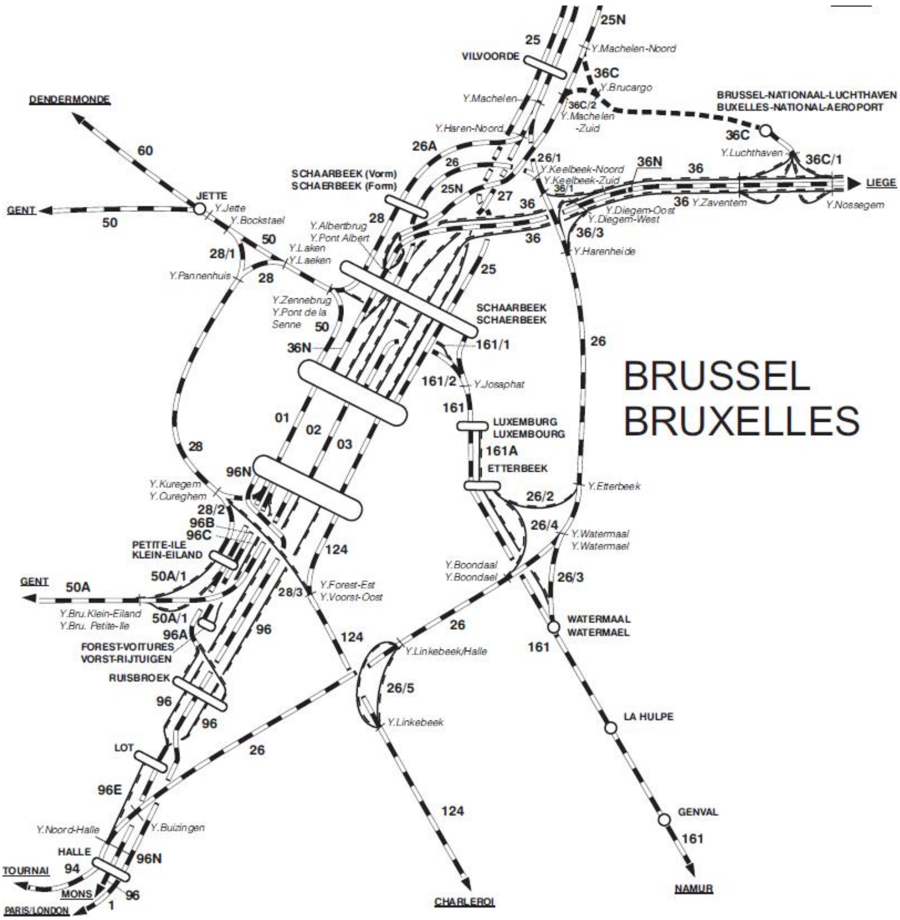


Figure 3.4: A schematic overview of the entire station area of Brussels. The tracks indicated by 01, 02 and 03 are a simplification of the core of Brussels dense railway area presented in Figure 3.3 (Dewilde, 2014b). Detailed track plans for Brussels can be found at http://www.sporenplan.nl/.

their route in the platform area of Brussels-South and are re-used for a second trip starting from this platform area. All trains dwell in Brussels-South, 76 in Brussels-Central, 76 in Brussels-North, 23 in Brussels-Schaarbeek, ten in Brussels-Congres and two in Brussels-Kapel.

3.6.3 Blocking times

Here the data is described that is used to calculate the lengths of the blocking times. The blocking time depends on the type of the train, on the route that the train is assigned to and the specific section on this route. Note that the release time of a section and the reservation time of the succeeding section are not equal, since the blocking time includes some safety time before the train actually enters the next section and there is some time that the tail of the train is still in the previous section while the head of the train is already in the next section. In this dissertation, the time to traverse a section is approximated by dividing the length of the section by the speed of the train in that section and the time to leave that section. If the train enters a section in which it stops at a platform, the minimum necessary dwell time for boarding and alighting of passengers (one up to five minutes) is added up with some additional time for braking to standstill and to accelerate back to the speed limit (up to 0.9 minutes). In consultation with the Belgian railway infrastructure manager Infrabel, we do not consider the possibility that a train needs multiple sections to brake because of the moderate train speeds in Brussels station area. In this chapter, the amount of additional time for braking and accelerating is based on data from Infrabel and depends on the type of the train, the direction of the train and the station in which the train dwells.

By definition, all nodes that are part of the same section are reserved and released at the same time. Only in case the section contains a platform, the nodes before the platform are already released from the moment the train has arrived on the platform. This design decision is also what Infrabel uses when making the planning for (complex) station areas. Obviously, during operations, more nodes could be released as soon as the train has passed these nodes.

We distinguish between two kinds of supplements that can be assigned: dwell time supplements and running time supplements. A dwell time supplement of 7% means that the blocking time of the section that contains the train's platform now contains 1.07 times the technical minimum dwell time of that train in that station instead of just the technical minimum dwell time. If the running time supplement in a grid zone is 7%, then the running time supplement provided to a section in this grid zone, is 7% of the technical minimum running time of that section.

If a train leaves the network in one of the platform areas, the blocking time of the platform is lengthened with four extra minutes to account for the time that the train needs to free that platform. Also if a train starts its trip in one of the platform areas, extra time is provided in the blocking time of the platform. The time to clear and unlock a section is approximated by a fixed amount of time, independent of the train and the section. These assumptions correspond to data used by Infrabel. Since the available data does not explicitly contain values for the time for setting up, watching the signal, approaching the section and releasing the signal, we assume that these times are included in the available data. Due to this uncertainty in the data, we cannot assure conflict-freeness at the microscopic level for our case study. However, we can assure that the same data and principles to calculate the lengths of the blocking times are used by the Infrabel and Dewilde (2014b) when planning complex station areas. So it still makes sense to compare the resulting routing plans and timetables.

3.6.4 Generalizability

It is meaningful and interesting to compare the performance of the routing plans and timetables constructed in this dissertation to the performance of a timetable and routing plan from practice and the best routing plan and timetable found in the literature for Brussels railway station area. For the real world schedule, a timetable and routing plan from Infrabel are used that dates back to 2010. For the best schedule for the Brussels station area from the literature, the timetable and routing plan constructed by Dewilde (2014b) are used. For these timetables and routing plans, Infrabel and Dewilde (2014b) used the same data and principles to calculate the lengths of the blocking times as described in the previous section. Also the line plan coincides with the data from Infrabel and corresponds to the line plan used by Dewilde (2014b). The reference timetable of Dewilde (2014b) contains the same dwell and running times and the same (dwell and running time) supplements as the reference timetable from Infrabel, about 203 minutes in total.

In this dissertation, results presented for the case study of Brussels are always compared to the results of the best timetable and routing plan constructed with the approach of Dewilde (2014b) and the results of the reference timetable and routing plan from Infrabel. The performance of the different timetables and routing plans for Brussels under a specific delay scenario is always determined with the simulation tool of Dewilde (2014b), explained earlier in this dissertation. Throughout this dissertation, different delay scenarios are used to compare timetables and routing plans. To be clear, simulation results presented in the same table, are always obtained by simulating the same delay scenario. Results in different tables can be obtained by simulating different delay scenarios. Therefore, the delay scenario will always be clarified. Although Dewilde et al. (2014a) tested different delay scenarios with small input delays and observed that these gave comparable results, this is not investigated for the methodologies in this dissertation.

The timetable and routing plan used as input for the approach presented in this chapter (existing routing plan and timetable), are the 2010 reference timetable and routing plan from Infrabel. The same reference schedule from Infrabel was used as input by Dewilde (2014b). Since the approach described in this chapter does not alter the supplements, the timetables constructed in this chapter contain the same amount of dwell and running time as the reference timetable from Infrabel and from Dewilde (2014b).

The operation costs of the different timetables and routing plans presented in this dissertation are assumed to be similar. First, only the bottleneck of the network is taken into account and not the whole Belgian railway network. Secondly, only the routing plan and the arrival and departure times within one period are changed. This is in particular in contrast with the infrastructure cost to build an extra track/platform to improve the robustness of the railway system. Generally, it is assumed that extra travel time entails extra operator costs. Since a passenger robust schedule minimizes the total travel time in practice of passengers, which is positively correlated with the total train travel time, the effect on the operator cost is assumed to be small. The cost of the included supplements is compensated by the reduction in delays. Furthermore, the lines that must be scheduled are the same in each constructed timetable, so the 'line cost' remains the same over all timetables. Moreover, the coupling and re-usage requirements (in the bottleneck) are also taken into account for all timetables.

3.6.5 Delay scenario

This chapter assumes that each train enters the network with a delay drawn from an exponential distribution with an average in the interval [0 min, 8 min]. These averages can be different for each train and are fixed in advance. There are no assumptions made on the correlation between late trains and heavily loaded passenger trains. For each train it holds that its average value of the exponential distribution, used during simulation, is the same as its mean recurring delay, used during optimization. No dwell delays are imposed, nor other source delays within the boundaries of the considered network. As explained above, the results that are presented in this chapter are all obtained by using the same delay scenario, i.e. that all trains enter the network with a certain delay. It should be noted, however, that for this reason, the results for the approach of Dewilde (2014b) that we present in this chapter, slightly differ from those reported in (Dewilde, 2014b), as different delay scenarios were used in that research. In Dewilde (2014b) at most three quarter of the trains were assumed to be delayed when entering the network.

3.7 Results

In this section the results of the approach that takes passenger numbers and/or recurring delays into account are presented. The presented approach is implemented in C++ and ran on a desktop computer Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz. The model (3.4)-(3.8) is solved by using CPLEX 12.6. Tables 3.2 and 3.3 present an overview of the simulation results. As mentioned earlier, these results are the output of a simulation tool that performs 10 000 simulation runs in which for each train a delay is drawn from an exponential distribution. The optimization algorithm takes, for each of the considered approaches, 60 to 90 minutes and the simulation model takes a few seconds.

Table 3.2: Overview of the simulation results: the new approaches improve the passenger robustness up to 11.33%.

Timetable	Passenger robustness ($\cdot 10^6$ min)	Impro (%)	Tot. knock-on delay (min)	Impro (%)	% extra delayed trains
Ref.	4.15		211.53		47.71
Dewilde et al.	3.84	7.47	145.31	31.30	37.36
$P_{t,t'}$	3.77	9.16	142.17	32.79	38.53
$D_{t,t'}$	3.72	10.36	136.38	35.53	36.98
$P_{t,t'}D_{t,t'}$	3.68	11.33	138.40	34.57	37.66

The rows in Table 3.2 refer to the different methods used to construct the timetable and routing plan. The label *'Ref.'* refers to the reference timetable and routing plan from 2010 from the Belgian railway infrastructure manager Infrabel, that is used as input for the here presented approach and for the approach of Dewilde (2014b). The label *'Dewilde et al.'* refers to the timetable and routing plan constructed with the approach described in Dewilde (2014b). The label '$P_{t,t'}$' refers to the timetable and routing plan constructed with the approach that takes passenger numbers into account, as described in Section 3.3. The label '$D_{t,t'}$' refers to the timetable and routing plan constructed with the approach that takes recurring delays into account, as described in Section 3.4. The label '$P_{t,t'}D_{t,t'}$' refers to the timetable and routing plan constructed with the approach that takes both passenger numbers and recurring delays into account. The weights in the latter approach consist of the product of the affected passengers and the weights associated with the recurring delays of both trains.

Since different weights are used in the different approaches, it doesn't make sense to compare the objective function values. The first column in Table 3.2

presents the passenger robustness value, averaged over the simulation runs. The lower this value, the better it is, as it represents the total weighted travel time in practice of all passengers. The approaches taking passenger numbers and/or recurring delays into account further improve the passenger robustness by $1.69\% - 3.86\%$ compared to the approach of Dewilde (2014b). In comparison with the reference timetable and routing plan from Infrabel, the passenger robustness improves with $9.16\% - 11.33\%$. The third column presents the total knock-on delay of all trains in one hour, expressed in train minutes. This is the average value over all simulation runs. The lower this value, the less the trains suffer from knock-on delays. Also here, the new approaches give an improvement of $1.49\% - 4.23\%$ compared with the results from Dewilde (2014b). The last column presents the percentage of trains that leave the network with a larger delay than that it entered the network. The number of extra delayed trains slightly deteriorates in the approaches where passenger numbers are taken into account compared to the approach of Dewilde (2014b). This is easily understandable, since the objective function in the optimization approach of Dewilde (2014b) focuses on trains (and is not specified in terms of 'passengers'). Maximizing the spreading between trains causes the number of trains that gets delayed to diminish. However, it could be more profitable to delay a number of trains with few passengers than one train with many passengers. The approach that takes recurring delays into account, by contrast, reduces the number of extra delayed trains compared to the approach of Dewilde (2014b).

Table 3.3: Overview of the simulation results when taking both recurring delays and passenger numbers into account during optimization.

Timetable	Passenger robustness ($\cdot 10^6$ min)	Impro (%)	Tot. knock-on delay (min)	Impro (%)	% extra delayed trains
Ref.	4.15		211.53		47.71
Dewilde et al.	3.84	7.47	145.31	31.30	37.36
$P_{t,t'} D_{t,t'}$	3.68	11.33	138.40	34.57	37.66
$1^{\text{st}} : P_{t,t'}$, $2^{\text{nd}} : D_{t,t'}$	3.71	10.60	133.66	36.81	37.13
$1^{\text{st}} : D_{t,t'}$, $2^{\text{nd}} : P_{t,t'}$	3.74	9.88	134.04	36.63	36.64

Table 3.3 presents the effect of two optimization rounds, where one optimization round refers to the completion of the entire optimization procedure, i.e. all iterations. The input of the second optimization round is the output of the first optimization round. The first rows with labels *'Ref'*, *'Dewilde et al.'* and '$P_{t,t'} D_{t,t'}$' coincide with the eponymous rows in Table 3.2. The row with label '$1^{st} : P_{t,t'}, 2^{nd} : D_{t,t'}$' uses the weights with the passenger numbers in the first round and the weights with the recurring delays in the second round. So,

recurring delays are not taken into account in the first round and passenger numbers are not taken into account in the second round. For the row with label '$1^{st}: D_{t,t'}, 2^{nd}: P_{t,t'}$' it is the other way round. Both approaches improve the passenger robustness slightly compared to the value when only passenger numbers or recurring delays are taken into account. The total knock-on delay per hour largely improves and there is a small improvement in the number of extra delayed trains. However, the combination of both weights in one optimization round remains the best option to improve the passenger robustness.

3.8 Conclusion

In this chapter, passenger numbers and recurring delays are taken into account to improve the passenger robustness in railway bottlenecks. The method presented in this chapter extends the algorithm developed by Dewilde et al. (2013, 2014a) and Dewilde (2014b). This algorithm optimizes the passenger robustness of an existing timetable and routing plan by improving the spreading of trains in a fixed time window. Passenger numbers and recurring delays are taken into account by weighing the costs associated with the available buffer time between a train pair. The more passengers can be affected by a propagated delay, the more important the corresponding buffer time. The larger the mean recurring delay of the first train of a train pair, the more important the corresponding buffer time becomes and the larger the mean recurring delay of the second train, the less important the corresponding buffer time becomes. These weights are included in the objective function and play a role in the selection of train pairs that are considered for increasing the buffer time. Prioritizing based on passenger numbers and recurring delays can be combined and proves to be successful. Computational results for Brussels railway area, the largest bottleneck in the Belgian railway network, show that the proposed extension further improves the passenger robustness up to 11.33%, which is in itself an improvement of 3.86% compared to the results of Dewilde et al. (2013, 2014a), Dewilde (2014b).

3.9 Future research

Further on in this thesis, more structural changes will be made to the algorithm presented here. The idea to achieve a passenger robust schedule by maximizing the appropriate buffer times between trains, remains the same. The new algorithm, however, is able to construct a passenger robust timetable and routing plan from scratch and even includes decisions on the amount and location of supplements.

An idea for further research is to include recurring *dwell* delays in the optimization model in addition to the recurring *arrival* delays. Another idea is to use this approach to improve passenger transfers. In fact, this is technically already included in the algorithm, but until now, no data is available to investigate the effect of passenger transfers in Brussels on the passenger robustness. Specific transfer times for a list of transfers could be taken into account. From the simulation output, information on the contribution of these transfer times on the passenger travel time in practice in combination with the delays of the involved trains and the missed transfers can be gathered. This information can then be used to improve the planning. It is also interesting to check how sensitive the approach is concerning the passenger numbers and the recurring delays that are used as input.

4

Robust routing and timetabling in complex railway stations

The focus of this chapter is also on complex and dense station areas.[1] Spreading the trains well in time and space in these areas has a big impact on passenger robustness. The aim is to improve the performance in the bottleneck of the network in order to improve the performance of the whole railway network. This chapter presents a method that builds a routing plan and a cyclic timetable *from scratch* that optimize the infrastructure occupation and the passenger robustness. Without considering a timetable, an integer linear routing model assigns every train to a route, such that the maximal node usage is minimized and that the number of times that each node is used, is quadratically penalized. Thereafter, a mixed integer linear timetabling model assigns to each train the blocking times at which the nodes on the train's route, assigned by the routing model, are reserved and released. The difference with other approaches is that the focus is on the occupation of the railway infrastructure, before constructing the timetable.

[1] This chapter brings together the peer-reviewed journal papers:

Sofie Burggraeve and Pieter Vansteenwegen. Robust routing and timetabling in complex railway stations. *Transportation Research Part B: Methodology*, 101:228–244, 2017c.

and:

Sofie Burggraeve and Pieter Vansteenwegen. Optimization of supplements and buffer times in passenger robust timetabling. *Journal of Rail Transport Planning & Management*, In Press, 2017b. URL `https://doi.org/10.1016/j.jrtpm.2017.08.004`.

Most sentences are cited from these papers and some are cited from Burggraeve et al. (2015c), but in order to keep the text easily readable, cited sentences are not separately indicated.

Furthermore, this chapter includes the routing and the timetabling model in an iterative approach that also takes supplement allocation into account on the signaling level. Each iteration is based on four pillars, which are subsequently executed. The first pillar consists of the routing model. The second pillar consists of the timetabling model. In the third pillar, a simulation is run to evaluate the passenger robustness of the routing plan and the timetable. The fourth pillar uses the simulation outcome to make a new supplement assignment for the next iteration.

The approach is validated on the complex railway station area of Brussels (Belgium). The passenger robustness improves up to 17% compared to a reference timetable and routing plan composed by the Belgian railway infrastructure manager Infrabel and by up to 8% compared to a timetable and routing plan developed by Dewilde et al. (2013, 2014a) and Dewilde (2014b) in exactly the same set-up.

4.1 Introduction

If a number of stations close to each other serves a high proportion of the passengers, this station area typically becomes a bottleneck in the railway system. A conflict in the railway bottleneck may affect the travel times of many passengers, since the conflict easily propagates delays to other trains driving through the bottleneck. So, the aim is to construct a conflict-free railway schedule for which the passenger robustness is optimized, i.e. the passenger travel time in practice, in case of frequently occurring small delays, is minimized. Unfortunately, direct implementation of this objective function is computationally highly demanding, since real travel times of all passengers and propagation of delays have to be calculated. Therefore passenger robustness is indirectly striven for by looking for a solution which optimally spreads the trains in time and space. This chapter is restricted to research on timetabling and railway routing, which are situated on the tactical level of railway planning. However, also line planning, on the strategic level, and real-time interventions, on the operational level, have an impact on the travel times of railway passengers in practice and thus on the passenger robustness of the railway system. Nevertheless, these stages are not considered here. Furthermore, the timetable and routing plan construction are only designed to mitigate the effect of frequently occurring small delays on the passenger travel times. The impact of large disturbances is not considered during the construction of timetables and routing plans. Nevertheless, we are convinced that this optimal spreading can also be useful in case of larger disturbances. To summarize, the focus of this research is on making a conflict-free and passenger robust timetable and a routing plan from scratch to transport passengers as passenger robust as possible into, through and out of a railway bottleneck.

The main contributions of this chapter are

- A routing model and a timetabling model to construct a conflict-free and passenger robust microscopic routing plan and timetable from scratch.
- An iterative approach including the routing and the timetabling model and a feedback loop that helps constructing even more passenger robust schedules by considering the inclusion of supplements as an extra degree of freedom.
- A validation of the routing model, the timetabling model and the iterative approach on a realistic, large and complex railway bottleneck, namely Brussels (Belgium).
- Passenger robust routing plans and timetables, built from scratch, for Brussels which significantly improve the passenger robustness of reference routing plans and timetables from practice and from literature.
- Additive and alternative constraints to speed up routing and timetabling models and to include transfers, re-usage, splitting and coupling of trains into these models.

The proposed routing and timetabling model can be used in sequence as illustrated in this chapter. However, they can also be used independently from each other.

The iterative approach can be seen as an extension of the routing and the timetabling model. This extended approach can be summarized as follows. Starting from a given assignment of supplements, the buffer times between trains are optimized in the construction of a conflict-free microscopic routing plan and timetable. The routing plan and the timetable are simulated and information on the 'causers' and the 'victims' of delay propagation are used to construct a new detailed assignment of supplements. This new assignment of supplements is the input for the next optimization round. Clearly, including more supplements reduces the room available for buffer times. The optimization stops when a balance between the achieved buffer times and the updated supplements is found that results in a passenger robust timetable.

In Section 4.2 the routing model and the timetabling model are explained in detail. The optimization models are introduced and illustrated within the context of a small case study example. The realistic case study is specified in Section 4.3. In Section 4.4, alternative and additional constraints are introduced in order to model the splitting, coupling and re-usage of trains, to include passenger transfers and to speed up the models. The performance of the routing and timetabling model is discussed in Section 4.5. Section 4.6 elaborates on the extended approach. The case study for the extended approach is specified in Section 4.7. The results of the extended approach are shown in Section 4.8. Conclusions and ideas for future research are presented in Section 4.9 and Section 4.10 respectively.

4.2 Methodology

This section introduces a new routing and timetabling model. The routing plan is constructed before the timetable is constructed. An advantage of this planning order is that all route assignments are still possible. No timetable is restricting the routing possibilities. The trains can be freely spread over the network in a way that is optimal for the infrastructure usage and the spreading of the trains in space. Another advantage of (microscopic) railway routing before (microscopic) timetabling is that one avoids in this way that no feasible routing through a complex station exists for a macroscopic timetable, as occurs for approaches that reverse both planning phases. This inconvenience is for example encountered by Sels (2016b). A disadvantage is that the existence of a feasible timetable for the constructed routing plan is not assured. If the railway routing and the timetabling are in a first phase restricted to the bottleneck of the network, then also a feasible extension of the planning to the remainder of the network is not assured. For the latter, however, Caimi et al. (2009a,b) booked good results for the Swiss railway network. This is not experimentally tested for the Belgian railway network. A roadmap to design a timetable and routing plan for a railway network with a bottleneck can be as follows:

1. First the network is divided in a bottleneck and the remainder of the network.
2. A microscopic routing is constructed to route the dense traffic over the complex infrastructure lay-out of the bottleneck, such that the infrastructure is optimally used.
3. A microscopic timetable for the bottleneck is constructed that maximizes the buffer times between the trains, where the amount of passengers on the trains is taken into account.
4. The timetable is extended to the remainder of the network.
5. Routing plans are constructed for in the station areas outside the bottleneck.

In the remainder of this section is focused on the construction of the routing plan and the timetable for the railway bottleneck. First the input and the output are described. Then the routing model is presented. The objective function, the parameters, the decision variables and the model itself are explained in detail. Thereafter the timetabling model is presented in the same way. This section ends with the illustration of the routing and the timetabling model on a small case study example.

4.2.1 Input

The input consists of the infrastructure for which a timetable and a routing plan are constructed. This includes the locations of the signals, the switches and the platforms. The routing model, that is executed first, starts from a fixed line plan that contains information on the incoming and outgoing tracks, the stops and the frequency of each train on the considered (part of the) network. The speed limits on all sections and the minimum dwell times are input to both the routing and the timetabling model. The speed limits are used to calculate the minimum running time of the trains based on the train type, e.g. high speed train, intercity train, etc. To accommodate for the speeding up and slowing down in the platform areas, fixed times, based on the platform area, the train direction and train type, are also considered as input and used for calculating the running times. On top of the minimum running time and dwell time, an initial supplement assignment can be used as input to be included in the running and dwell times.

4.2.2 Output

The output consists of the routing plan that spreads the trains best in space and the timetable that spreads the trains best in time. Both the output routing plan and the timetable are specified on the microscopic level.

4.2.3 Routing model

The routing model assigns every train to a route. A route is defined as a sequence of succeeding nodes and links. We focus on the nodes because they uniquely determine the route of a train through the network. Moreover, two trains can only be in conflict if they share at least one node. It is possible that trains don’t share a (complete) section but that they only share one node (as a cross point of two sections). Furthermore, the infrastructure occupation of a railway system can be expressed by the usage of the nodes in the network. We make no assumptions about the timetable as the timetable will only be determined after the routing plan. However, the routing plan is constructed such that a passenger robust timetable can be built. The routing model assigns every train to a route such that every node is used as little as possible. Consequently, the trains are spread in space and the infrastructure of the station area is optimally used. The underlying logic is that the less train traffic there is on a node, the easier it will be to spread individual trains over time during timetabling in order to construct a passenger robust timetable. Not the duration of each usage is taken into account, but the number of usages. This will be discussed further in the results section.

Objective function

In order to achieve that every node is used as little as possible, we combine two aspects. We explicitly minimize the maximum use of any node, i.e. we minimize the number of usages of the most used node. We say that a node is used x times if there are x trains whose route contains this node. Furthermore, we minimize the sum of the squared usages of all nodes. We minimize the sum of the squared usages, instead of for example the sum of the usages itself, in order to penalize an increase in a node utilization more heavily when the utilization rate of that node is already high. This second objective gives the incentive to further decrease the individual node usages. Intuitively, slalom routes pass more nodes and will thus negatively affect our objective function. These kind of routes will be avoided. The two minimization problems are integrated into one problem by giving the minimization of the maximum node usage a much higher weight ($\mathcal{H}$). The magnitude of $\mathcal{H}$ depends on the problem size. Now, the optimization model is presented and explained.

Parameters

$$
\begin{array}{lcl}
T & = & \{t_1, t_2, \cdots, t_{|T|}\} \\
 & = & \text{set of trains with } |T| \text{ the number of trains.} \\
W & = & \{w_1, w_2, \cdots, w_{|W|}\} \\
 & = & \text{set of nodes with } |W| \text{ the number of nodes.} \\
R & = & \{r_1, r_2, \cdots, r_{|R|}\} \\
 & = & \text{set of routes with } |R| \text{ the number of routes.} \\
R_t \subset R & = & \text{set of routes that train } t \text{ can be assigned to (based on train } t\text{'s} \\
 & & \text{incoming and outgoing line in the station area and its stops).}
\end{array}
$$

$l_{r,w}$	$=$	1 if route $r \in R$ contains node $w \in W$ and
	$=$	0 otherwise.
$\mathcal{H}$	$=$	the high weight used to enforce the domination of the minimax criterion in the objective function of the routing model.

Decision variables

$g_w \in \mathbb{Z}^+$	$=$	number of times node w is used, $w \in W$.
$x_{t,r}$	$=$	1 if train $t \in T$ is assigned to route $r \in R_t$ and
	$=$	0 otherwise.
z	$=$	overall maximum node usage.

Model

The combined weighted objective function of the minimization of the maximum node usage and the minimization of the sum of the squared node usages, can be represented by

$$\min \qquad \mathcal{H} \max_{w \in W} g_w + \sum_{w \in W} g_w^2. \tag{4.1}$$

Note that both terms of this objective function are non-linear. The first part is a minimax objective and can easily be linearized by introducing a decision variable z and constraints (4.6). The second part is harder to linearize as it consists of a sum of squared but bounded integers. A linearization method is now described, together with two different implementations. One of these implementations proved to result in a much shorter computation time in our case study.

Introduce $K|W|$ binary decision variables $b_{k,w}$, with $k \in \{1, \cdots, K\} \subset \mathbb{Z}^+, w \in W$ and with K an upper bound on the maximum node usage. A train can pass a node at most once, so in any case $K \leq |T|$. The $b_{k,w}$'s are defined as

$$b_{k,w} = \begin{cases} 1 & \text{if} \quad g_w \geq k \\ 0 & \text{if} \quad g_w \leq k-1 \end{cases} \tag{4.2}$$

If node w is used at least k times ($g_w \geq k$), then $b_{k,w} = 1$, else $b_{k,w} = 0$. The sum of squared node usages is converted to a sum of linear sums by using the coefficients c_k given in Table 4.1 for each node w:

$$\sum_{k=1}^{K} c_k b_{k,w} = \sum_{k=1}^{g_w} c_k = g_w^2, \tag{4.3}$$

where the first equality is true by definition of the $b_{k,w}$'s.

Table 4.1: Linearization of the quadratic terms.

	c_m	$\sum_{k=1}^{m} c_k$
$m \in \mathbb{Z}^+$	$2m-1$	m^2

Table 4.2 explains the values in the third column of Table 4.1.

Table 4.2: Linearization of the quadratic terms.

m	c_m	$\sum_{k=1}^{m} c_k$
1	1	1
2	3	$1+3=4$
3	5	$1+3+5=9$
$\vdots$	$\vdots$	$\vdots$
K	$2K-1$	$1+3+\cdots+2K-3+2K-1$
even		$=\sum_{\beta=2\alpha-1,\alpha\in\{1,2,\cdots,\frac{K}{2}\}}(\beta+2K-\beta)=\frac{K}{2}\cdot 2K=K^2$
odd		$=K+\sum_{\beta=2\alpha-1,\alpha\in\{1,2,\cdots,\frac{K-1}{2}\}}(\beta+2K-\beta)$
		$=K+\frac{K-1}{2}\cdot 2K=K^2$

A first option to define the $b_{k,w}$'s in this minimization problem is to use the following constraints:

$$g_w - Kb_{k,w} \leq k-1 \qquad \forall k \in \{1,\cdots,K\}, \forall w \in W. \tag{4.4}$$

A second option is to use constraints (4.7). Both implementations give rise to the same optimal value. However, the number of constraints in (4.7) is only $|W|$ compared to $K|W|$ in (4.4). The computation times of both implementations are compared in Section 4.5. Including constraints (4.7) clearly outperforms the inclusion of constraints (4.4). This results in the following integer linear routing model, where constraints (4.7) are included:

$$\min \quad \mathcal{H}z + \sum_{w\in W}\left(\sum_{k=1}^{K} c_k b_{k,w}\right) \tag{4.5}$$

$$\text{s.t.} \quad g_w \leq z \qquad \forall w \in W \tag{4.6}$$

$$g_w = \sum_{k=1}^{K} b_{k,w} \qquad \forall w \in W \tag{4.7}$$

$$\sum_{r\in R_t} x_{t,r} = 1 \qquad \forall t \in T \tag{4.8}$$

$$\sum_{t\in T}\sum_{r\in R_t} l_{r,w}x_{t,r} = g_w \qquad \forall w \in W \tag{4.9}$$

$$g_w \in \mathbb{Z}^+ \qquad \forall w \in W \tag{4.10}$$

$$z \in \mathbb{Z}^+ \tag{4.11}$$

$$x_{t,r} \in \{0,1\} \qquad \forall t \in T, \forall r \in R_t \tag{4.12}$$

$$b_{k,w} \in \{0,1\} \qquad \forall k \in \{1,\cdots,K\}, w \in W. \tag{4.13}$$

Constraints (4.8) assure that each train is assigned to exactly one route that connects the incoming and outgoing line of that train in the network. Constraints (4.9) assure that g_w equals the number of trains that will use node w. In fact, the decision variable g_w is not strictly necessary and can be omitted from the model, by everywhere replacing g_w with the expression in constraints (4.9). The g_w's are introduced into the model, because they represent the node usage, which is the main performance indicator of the routing model. Constraints (4.10) and (4.12) assure that the number of times that a node is used is a positive integer and that the assignment of a train to a route is represented by the binary variable $x_{t,r}$. If the $b_{k,w}$'s would only be defined as continuous variables in $[0,1]$, instead of binary variables in (4.13), and z as a continuous variable in $\mathbb{R}_0^+$, instead of a positive integer in (4.11), their values would still be binary, respectively integral, in the optimal solution due to the model construction.

4.2.4 Timetabling model

We propose a new mixed integer linear model to construct a cyclic timetable from scratch. The overall objective is to build a conflict-free and passenger robust timetable on microscopic level. The timetabling model uses a microscopic routing plan as input, for example the routing plan assigned by the routing model. The dwell and running times and the lengths of the blocking times of the sections on the train's route can be calculated beforehand based on the characteristics of the train, the section and the supplement distribution. These blocking time lengths are used to fix time durations that indicate the reservation and release times of the sections on the train's route relative to the entrance time of the train at the border node of its incoming line. These time durations are used as parameters in the timetabling model. For each train, its reservation and release times of nodes that are part of the same section, are the same.

Objective function

The objective function of the timetabling model is to maximize the buffer times between trains. Sufficient buffer times between trains avoid the propagation of delays and favor shorter and more reliable travel times in practice for passengers. The buffer times in a certain node between two trains (which trips are repeated every period) are presented on a cyclic time axis in Figure 4.1.

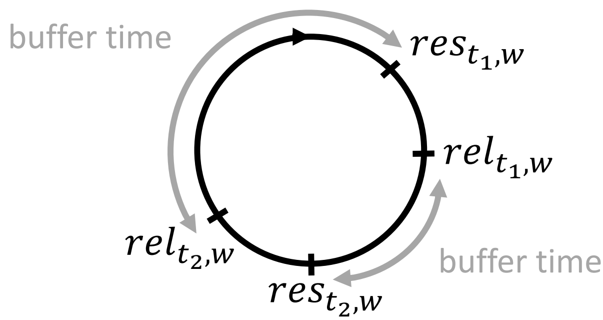


Figure 4.1: Train t_1 reserves and releases the node at time instants $\mathrm{res}_{t_1,w}$ and $\mathrm{rel}_{t_1,w}$. Train t_2 reserves and releases that same node at time instants $\mathrm{res}_{t_2,w}$ and $\mathrm{rel}_{t_2,w}$. A first buffer time is the one between $\mathrm{rel}_{t_1,w}$ and $\mathrm{res}_{t_2,w}$ and the second buffer time is the one between $\mathrm{rel}_{t_2,w}$ and $\mathrm{res}_{t_1,w}$.

To optimize the buffer times between the trains, we again focus on two aspects. The approach is similar to that of the routing model. Primarily, we maximize the minimum buffer time over all nodes, i.e. the smallest amount of time between the reserve and release times of trains on any node. Note that the total amount of buffer time available in a node is fixed if the number of passing trains and the lengths of their blocking times are fixed. The spreading of this total amount over different buffer times, by contrast, can be changed. This is visualized in Figure 4.2. Striving toward larger buffer times in a node thus coincides with striving toward buffer times with equal length. Striving toward buffer times of equal lengths is achieved by striving to maximize the minimum buffer time in a node. Secondarily, we maximize the sum of the minimum buffer times between every two trains having at least one node in common. This second optimization objective gives the incentive to further increase the buffer times between train pairs. We again combine these two objectives by adding them up and by giving a high weight ($\widetilde{\mathcal{H}}$), depending on the problem size, to the overall minimum buffer time.

No decisions on supplements are included in the routing and the timetabling

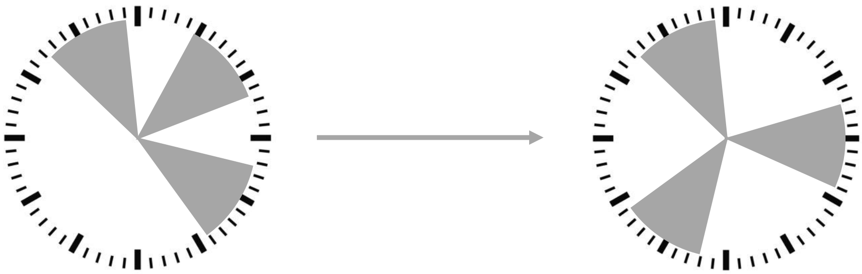


Figure 4.2: The total amount of buffer time (the white space between the gray triangles) is fixed but the spreading of the buffer times is variable.

model. However, different amounts of dwell and running time supplement may be used as input to the routing and the timetabling model. The results of different input amounts of supplements are shown in Section 4.5. The more supplements are added in the timetable, the more delays can be absorbed, but the less time and capacity is left for including buffer times between train pairs. Later in this chapter, this approach is extended to an iterative approach in which the inclusion of supplements acts as an extra degree of freedom.

Parameters

r^t	=	the route that is assigned to train $t \in T$ in the routing plan.
s_r (s_{r^t})	=	the first node on route r (r^t).
$d^{\text{res}}_{t,r^t,w}$ $(d^{\text{rel}}_{t,r^t,w})$	=	the time duration for train $t \in T$ to reserve (release) node w relatively to the entrance of train t in the network.
T_w	=	the set of trains t for which r^t contains node w.
P	=	the period length of the cyclic timetable.
$\widetilde{\mathcal{H}}$	=	the high weight used to impose the domination of the maximin criterion in the objective function of the timetabling model.

In the timetabling model, we make use of variables that indicate the reservation and release times of the trains in the cyclic timetable. So, the value of these variables is restricted to the interval $[0, P[$. Furthermore, we also make use of auxiliary variables for the reservation and auxiliary release times of the trains, which value can be outside the interval $[0, P[$, more specifically, these values can be larger than P, not smaller than zero. These auxiliary reservation and release variables are used to easily link the different events of one train to each other, where events must be interpreted as reservations and releases of sections and nodes on the trains' routes. Remember that once the routes are known, we exactly know how long a train will block each part on its route. Based on these blocking times, all the event times can be defined relatively to each other. This is done by using the auxiliary reservation and release times: the reservation time of a certain section or node is the reservation time of the first section plus the according time duration between these two events, the release time of a certain section or node is the reservation time of the first section plus the according time duration between these two events. Note now that if the reservation time of the first section is assigned the value $P-5$ and the time duration between this reservation time and the reservation time of a section later on the train's route is larger than five minutes, the sum of the reservation time of the first section and the time duration will exceed P. Therefore we assign this value to the auxiliary reservation time of that later section. Before the model is optimized, we do not know which value will be assigned to the reservation time of the first section on the train's route and thus we do not know whether (and if yes for which events) the period length will be exceeded. The relation between the reservation and release times and the corresponding auxiliary reservation and auxiliary release times is that their difference is always a multiple of the period length. We will refer to the set of all auxiliary reservation and auxiliary release times as the *auxiliary timetable* and the set of all reservation and release times as the *cyclic timetable*.

Decision variables

Let ξ be the smallest integer such that all $d^{\text{res}}_{t,r^t,w}$ and $d^{\text{rel}}_{t,r^t,w}$ are smaller than $(\xi-1)P$.

$\text{res}_{t,w}, \text{rel}_{t,w} \in [0, P[$	$=$	the moment at which train $t \in T$ reserves, resp. releases node w in the cyclic timetable, $\forall t \in T, w \in r^t$.
$\text{res}^{\text{aux}}_{t,w}, \text{rel}^{\text{aux}}_{t,w} \in [0, \xi P[$	$=$	the moment at which train $t \in T$ reserves, resp. releases node w in the auxiliary timetable, $\forall t \in T, w \in r^t$.
$\text{res}^{\text{int}}_{t,w}, \text{rel}^{\text{int}}_{t,w} \in \{0, 1, \cdots, \xi-1\}$	$=$	h if $\text{res}^{\text{aux}}_{t,w}$, resp. $\text{rel}^{\text{aux}}_{t,w} \in [hP, (h+1)P[$, $\forall t \in T, w \in r^t$, $h \in \{0, 1, \cdots, \xi-1\} \subset \mathbb{Z}^+$.

$\text{resrel}^{\text{int}}_{t_i,t_j,w} \in \{-\xi, \cdots, \xi\}$	$=$	h if $\text{res}^{\text{aux}}_{t_j,w} + hP - \text{rel}^{\text{aux}}_{t_i,w} \in [0, P[$, $\forall t_i, t_j \in T : i < j, w \in r^{t_i} \cap r^{t_j}$, $h \in \{-\xi, \cdots, \xi\} \subset \mathbb{Z}$.
$\text{buf}_{t_i,t_j,w} \in [0, P[$	$=$	the buffer time between the reservation time of train t_j, and the release time of train t_i in node w, $\forall t_i, t_j \in T : t_i \neq t_j, w \in r^{t_i} \cap r^{t_j}$.
$\text{minbuf}_{t_i,t_j} \in [0, P/2]$	$=$	the minimum buffer time between train t_i and t_j
	$=$	$\min\{\text{buf}_{t,t',w} \mid t, t' \in \{t_i, t_j\} : t \neq t', w \in r^{t_i} \cap r^{t_j}\}$ $\forall t_i, t_j \in T : t_i \neq t_j, r^{t_i} \cap r^{t_j} \neq \emptyset$.
$\tilde{z} \in [0, P/2]$	$=$	the overall minimum buffer time.

Model

The maximization of the minimum buffer time over all nodes can be represented by:

$$\max \quad \min\{\text{buf}_{t_i,t_j,w} \mid w \in W, (t_i, t_j) \in T_w \times T_w : t_i \neq t_j\}. \tag{4.14}$$

The maximization of the sum of the minimum buffer times between every two trains can be represented by:

$$\max \sum_{\substack{t_i,t_j \in T:\, i<j \\ \wedge\, r^{t_i} \cap r^{t_j} \neq \emptyset}} \min\{\text{buf}_{t,t',w} \mid t, t' \in \{t_i, t_j\} : t \neq t', w \in r^{t_i} \cap r^{t_j}\}. \tag{4.15}$$

Both parts of the objective function are or can be reduced to maximin objectives. Objective function (4.16) together with constraints (4.17) (for the first objective part) and constraints (4.18) (for the second objective part) present a linearization of the combined weighted objective function. The timetabling model then becomes:

$$\max \qquad \widetilde{\mathcal{H}}\tilde{z} + \sum_{\substack{t_i,t_j \in T:\, i<j \\ \wedge\, r^{t_i} \cap r^{t_j} \neq \emptyset}} \text{minbuf}_{t_i,t_j} \tag{4.16}$$

such that

$$\tilde{z} \leq \text{buf}_{t_i,t_j,w} \qquad \forall w \in W, t_i, t_j \in T_w : t_i \neq t_j \tag{4.17}$$

$$\text{minbuf}_{t_{\min\{i,j\}}, t_{\max\{i,j\}}} \leq \text{buf}_{t_i,t_j,w} \qquad \forall t_i, t_j \in T : t_i \neq t_j, w \in r^{t_i} \cap r^{t_j} \tag{4.18}$$

$$\mathrm{res}^{\mathrm{aux}}_{t_j,w} - \mathrm{rel}^{\mathrm{aux}}_{t_i,w} \geq -P\,\mathrm{resrel}^{\mathrm{int}}_{t_i,t_j,w} \qquad \forall w \in W, \forall t_i,t_j \in T_w : i < j \tag{4.19}$$

$$\mathrm{res}^{\mathrm{aux}}_{t_i,w} - \mathrm{rel}^{\mathrm{aux}}_{t_j,w} \geq -P(1-\mathrm{resrel}^{\mathrm{int}}_{t_i,t_j,w}) \qquad \forall w \in W, \forall t_i,t_j \in T_w : i < j \tag{4.20}$$

$$\mathrm{buf}_{t_i,t_j,w} \leq \mathrm{res}^{\mathrm{aux}}_{t_j,w} - \mathrm{rel}^{\mathrm{aux}}_{t_i,w} + P\,\mathrm{resrel}^{\mathrm{int}}_{t_i,t_j,w} \qquad \forall w \in W, \forall t_i,t_j \in T_w : i < j \tag{4.21}$$

$$\mathrm{buf}_{t_j,t_i,w} \leq \mathrm{res}^{\mathrm{aux}}_{t_i,w} - \mathrm{rel}^{\mathrm{aux}}_{t_j,w} + P(1-\mathrm{resrel}^{\mathrm{int}}_{t_i,t_j,w}) \qquad \forall w \in W, \forall t_i,t_j \in T_w : i < j \tag{4.22}$$

$$\mathrm{res}^{\mathrm{aux}}_{t,w} - \mathrm{res}_{t,s_{r^t}} = d^{\mathrm{res}}_{t,r^t,w} \qquad \forall t \in T, \forall w \in r^t \tag{4.23}$$

$$\mathrm{rel}^{\mathrm{aux}}_{t,w} - \mathrm{res}_{t,s_{r^t}} = d^{\mathrm{rel}}_{t,r^t,w} \qquad \forall t \in T, \forall w \in r^t \tag{4.24}$$

$$0 \leq \mathrm{res}^{\mathrm{aux}}_{t,w} - P\,\mathrm{res}^{\mathrm{int}}_{t,w} < P \qquad \forall w \in W, \forall t \in T_w \tag{4.25}$$

$$\mathrm{res}_{t,w} = \mathrm{res}^{\mathrm{aux}}_{t,w} - P\,\mathrm{res}^{\mathrm{int}}_{t,w} \qquad \forall w \in W, \forall t \in T_w \tag{4.26}$$

$$0 \leq \mathrm{rel}^{\mathrm{aux}}_{t,w} - P\,\mathrm{rel}^{\mathrm{int}}_{t,w} < P \qquad \forall w \in W, \forall t \in T_w \tag{4.27}$$

$$\mathrm{rel}_{t,w} = \mathrm{rel}^{\mathrm{aux}}_{t,w} - P\,\mathrm{rel}^{\mathrm{int}}_{t,w} \qquad \forall w \in W, \forall t \in T_w \tag{4.28}$$

$$\tilde{z} \in [0, P[\tag{4.29}$$

$$\mathrm{minbuf}_{t_i,t_j} \in [0, P/2] \qquad \forall t_i,t_j \in T : i < j \tag{4.30}$$

$$\mathrm{buf}_{t_i,t_j,w} \in [0, P[\qquad \forall w \in W, \forall t_i,t_j \in T_w : t_i \neq t_j \tag{4.31}$$

$$\mathrm{res}_{t_i,w}, \mathrm{rel}_{t_i,w} \in [0, P[\qquad \forall w \in W, \forall t_i \in T_w \tag{4.32}$$

$$\mathrm{res}^{\mathrm{aux}}_{t_i,w}, \mathrm{rel}^{\mathrm{aux}}_{t_i,w} \in [0, \xi P[\qquad \forall w \in W, \forall t_i \in T_w \tag{4.33}$$

$$\mathrm{res}^{\mathrm{int}}_{t_i,w}, \mathrm{rel}^{\mathrm{int}}_{t_i,w} \in \{0, \cdots, \xi - 1\} \subset \mathbb{Z} \qquad \forall w \in W, \forall t_i \in T_w \tag{4.34}$$

$$\mathrm{resrel}^{\mathrm{int}}_{t_i,t_j,w} \in \{-\xi, \cdots, \xi\} \subset \mathbb{Z} \qquad \forall w \in W, \forall t_i,t_j \in T_w : i < j. \tag{4.35}$$

Constraints (4.19) - (4.20) assure that the blocking times of two trains that share a node do not overlap. These constraints need some extra explanation. In a non-cyclic timetable, blocking times of a train pair (t_i, t_j) for a certain node w do not overlap if the reservation time of one of the two trains is later than the release time of the other of the two trains, i.e. $\text{res}_{t_i,w} - \text{rel}_{t_j,w} \geq 0$ or $\text{res}_{t_j,w} - \text{rel}_{t_i,w} \geq 0$. However, in a cyclic timetable, this condition can be fulfilled for the auxiliary reservation and release times or for the reservation and release times itself while the blocking times do overlap. An example of each is provided in Table 4.5, considering that the period length P equals 60 minutes. A negative result in line 1 and 2 or 3 and 4 for the last two columns indicates an overlap of blocking times. Constraints (4.19) - (4.20) exclude these cases and force that the blocking times of different trains for a shared node cannot overlap.

Table 4.5: Example of (non)-straightforward overlapping blocking times.

Train	i	$(\text{res}^{aux}_{t_i,w}, \text{rel}^{aux}_{t_i,w}) \rightarrow (\text{res}_{t_i,w}, \text{rel}_{t_i,w})$			(i,j)	$\text{res}^{\text{aux}}_{t_i,w} - \text{rel}^{\text{aux}}_{t_j,w}$	$\text{res}_{t_i,w} - \text{rel}_{t_j,w}$
t_1	1	(2,5)	$\rightarrow$	(2,5)	(1,2)	2 - 66 = -64	2 - 6 = -4
t_2	2	(63,66)	$\rightarrow$	(3,6)	(2,1)	63 - 5 = 58	3 - 5 = -2
t_3	3	(58,61)	$\rightarrow$	(58,1)	(3,4)	58 - 62 = -4	58 - 2 = 56
t_4	4	(59,62)	$\rightarrow$	(59,2)	(4,3)	59 - 61 = -2	59 - 1 = 58

The auxiliary variable $\text{resrel}^{\text{int}}_{t_i,t_j,w}$ $(i < j)$ is a measure of how many times the period length has to be added to the auxiliary release time of train t_j such that the buffer time in node w, $\text{buf}_{t_i,t_j,w}$, can be calculated as the auxiliary release time of train t_j plus $\text{resrel}^{\text{int}}_{t_i,t_j,w}$ times the period length minus the auxiliary reservation time of train t_i. The buffer time between two trains in a shared node is defined by constraints (4.21)-(4.22). Constraints (4.23)-(4.24) link the reservation time of the start node to the reservation and release times of the other nodes on the route of a train, based on the duration parameters. Constraints (4.25)-(4.28) link the auxiliary reservation and release times to the (cyclic) reservation and release times. Constraints (4.29)-(4.35) are the type settings for the decision variables.

Note that the decision variables $\text{res}_{t,w}$ and $\text{rel}_{t,w}$ for all $w \in W$ and $t \in T_w$ are not strictly necessary in the model. In constraints (4.23) and (4.24), the variables $\text{res}_{t,s_{rt}}$ can be replaced by their auxiliary counterpart for all $t \in T$. Furthermore, constraints (4.25) - (4.28) can be removed, but extra constraints must be added that require that for all $t \in T$ the auxiliary reservation time of the first node of route r^t, $\text{res}^{aux}_{t,s_{rt}}$, belongs to the interval $[0, P[$. In the current representation of the model, it is only required that these latter variables belong to the interval $[0, \xi P[$. Since these $\text{res}_{t,w}$ and $\text{rel}_{t,w}$ variables are the variables that determine the cyclic timetable, the model formulation that includes these variables is preferred in this dissertation.

4.2.5 Small case study example

We demonstrate our routing and timetabling approach on the fictive and schematic network in Figure 4.3.

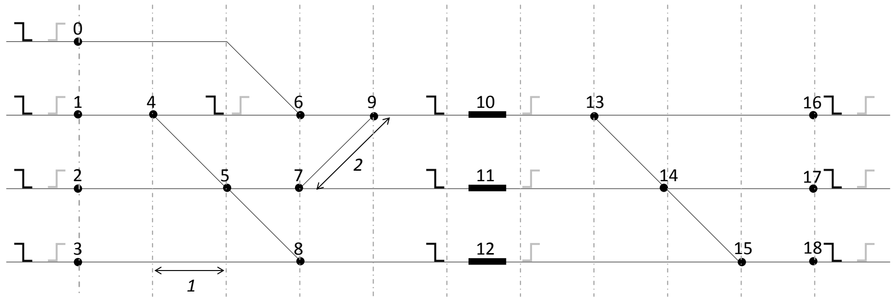


Figure 4.3: Small network example.

The bold circles represent switches or border nodes, the rectangles are platforms and the flags are signals which demarcate the sections on the network. Signals indicated in gray (black) are for trains driving from left (right) to right (left). Instead of calculating detailed blocking times, we use simple calculations to fix the time duration parameters. The numbers next to the arrows present the time (in minutes) necessary to get from one end of the corresponding link to the other end. For every slanted edge, this is two minutes, for every (part of a) horizontal edge, this is one minute. Edge parts are indicated with the crossings of vertical dashed lines. We suppose that a train needs 0.2 minutes to get its tail out of a section once its head left that section. The necessary dwell time at the station is two minutes.

We consider six trains on our network for which the incoming and outgoing tracks are fixed in advance. The origin and destination of each of these trains is indicated in Table 4.6. This table also lists the possible train routes for each train, where the numbers refer to Figure 4.4, which visualizes these routes. Each train has up to four possible routes.

The output of the routing model and the timetabling model is visualized in Figure 4.5. The maximum number of times that a node is used is three (node 5). The sum of the squared node usages is 75. Table 4.7 presents the reservation and release times of each train in all the nodes on its route. The nodes are put together per section. Table 4.8 shows the minimum buffer times between each train pair sharing at least one node. The overall minimum buffer time is 12.13 minutes, which occurs between the reservation and release times of trains 0, 3

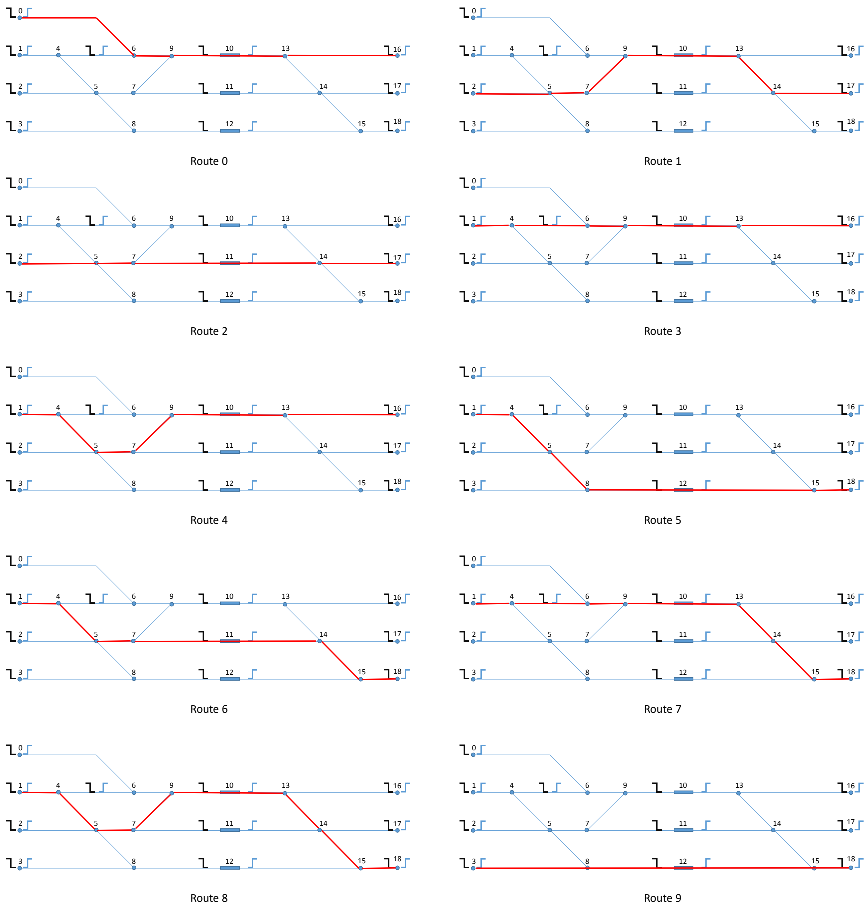


Figure 4.4: Possible train routes for trains 0 to 5 in Table 4.6.

and 4 in node 5. Note that node 5 is blocked for 23.6 minutes by these three trains, so the buffer times are equally spread in this node ($\frac{60-23.6}{3} = 12.13$). The sum of the minimum buffer times between every train pair sharing at least one node is 102.79 minutes.

Figure 4.6 shows an alternative optimal outcome of the routing model when only the maximum node usage is minimized. So here only the first part of the objective function (4.1) is taken into account. This shows that using also the second part of objective function (4.1) results in the fact that only one node is used three times instead of several nodes.

Table 4.6: Train lines and possible routes.

Train	Color	Origin	Destination	Possible routes
0	●	18	1	5, 6, 7, 8
1	●	18	3	9
2	●	16	1	3, 4
3	●	2	17	1, 2
4	●	2	17	1, 2
5	●	0	16	0

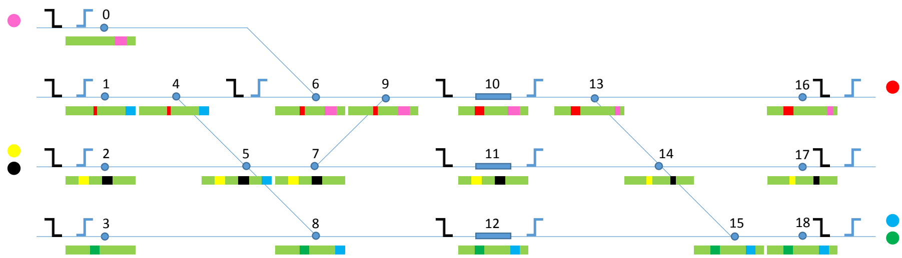


Figure 4.5: For each of the six trains a colored dot indicates the border node of its incoming track. The small rectangles below the nodes present a time axis of one hour indicating when the nodes are free (buffer times) and blocked. Parts indicated in light green are buffer times and the other colors indicate the blocking times of the different trains.

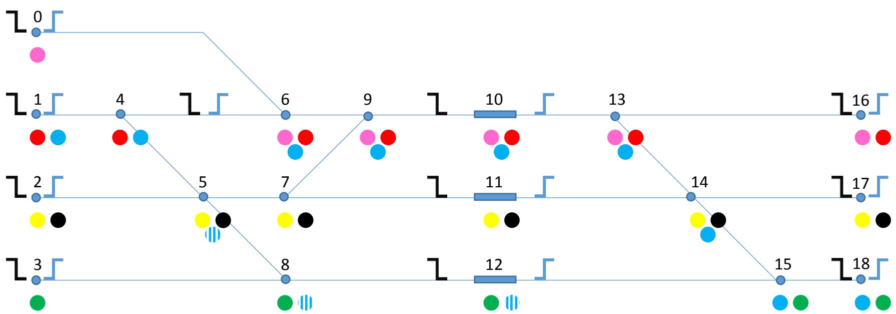


Figure 4.6: An alternative optimal routing plan when only the maximum node usage is minimized, i.e. only the first part of the objective function (4.1) is considered. Only the route of train 0 is different. This alternative solution assigns train 0 to route 7, which is indicated with solid light blue dots. The dots with a light blue hatched fill pattern indicate route 5, which proves better in case also the second part of the objective function (4.1) is taken into account.

Table 4.7: The reservation and release times for each train in all the nodes on its route.

Train t	Nodes w	$\text{res}_{t,w}$	$\text{rel}_{t,w}$
0	18; 15; 12	45.80	53.00
	8; 5; 4; 1	52.80	0.00
1	18; 15; 12	15.80	23.00
	8; 3	22.80	30.00
2	16; 13; 10	15.30	22.50
	9; 6	22.30	25.50
	4; 1	25.30	27.50
3	2; 5; 7; 11	12.13	20.33
	14; 17	20.13	24.33
4	2; 5; 7; 11	32.47	40.67
	14; 17	40.47	44.67
5	0; 6; 9; 10	43.80	53.00
	13; 16	52.80	57.00

Table 4.8: The overall minimum buffer time is 12.13 minutes. The sum of the minimum buffer times between each pair that has at least one node in common is 102.79 minutes.

Train pair	Minimum buffer time
0 - 1	22.80 min
0 - 2	25.30 min
0 - 3	12.13 min
0 - 4	12.13 min
2 - 5	18.30 min
3 - 4	12.13 min
Sum	102.79 min

4.3 Case study

To validate the proposed models, we use the case study that is described in Section 3.6. The blocking times are used to fix the time duration parameters $d^{\text{res}}_{t,r^t,w}$ and $d^{\text{rel}}_{t,r^t,w}$. In this chapter, during optimization, the blocking time of the first section of a train is lengthened with one minute by advancing the reservation time. This is to account for the approach time of the train towards the considered restricted network. Note that the presented timetabling model can be applied to any (part of a) network by appropriately assigning values to these duration parameters $d^{\text{res}}_{t,r^t,w}$and $d^{\text{rel}}_{t,r^t,w}$. This makes the timetabling model generally applicable; the only limit is the size and complexity of the network considered. Nevertheless, we show here that good results are obtained for the case study of Brussels, characterized by 481 nodes, 421 signals and 85 trains per hour. However, no tests are performed to construct a microscopic timetable for the whole Belgian railway network.

In order to rule out the effect on the passenger robustness of a different amount of supplements included in the schedule, the amount of dwell and running time supplements added in this chapter equals the amount of supplements present in the timetable of Dewilde (2014b) and the Belgian railway infrastructure manager Infrabel, i.e. about 203 minutes. These 203 minutes are all included as dwell time supplements and are added on top of the minimum necessary dwell time on a platform.

Now, alternative and additional constraints are presented in order to (i) speed up the models, (ii) model the re-usage, splitting and coupling of trains and (iii) model transfers. Thereafter the performance results will be presented and discussed.

4.4 Additional and alternative constraints for routing and timetabling models

In this section, additional and alternative constraints for routing and timetabling models are proposed that improve the construction and/or computation time of the models for large and dense networks or extend the practical applicability of these models. First some flow conservation constraints are introduced in the routing model in order to restrict the model construction time and the time of preliminary assembling routes. Then, some additional constraints are added to speed up the timetabling model. Thereafter, constraints to model splitting, coupling and re-usage of trains are introduced. Finally, extra constraints to include passenger transfers in routing and timetabling models are constructed.

4.4.1 Flow conservation constraints to speed up the routing model

The total number of routes starting at a border node or platform and ending at a border node or platform amounts to 28 970 248 for the Brussels station area. As input, the routing model needs for each train the set of routes that connects the border node or platform of its incoming and outgoing line. For each of these train-route combinations a variable has to be included in the model, which has to be considered in constraints (4.9) for each node on that route. To decrease the computational effort, we divide the network in four parts to construct the routes. These parts are determined based on the most important platform areas. We call a route that is restricted to one part of the network a *partial route*. The four parts are indicated in Table 4.9, together with their number of partial routes in each direction. This division in partial routes significantly reduces the terms in constraints (4.9). This reduction can be explained as follows. Suppose that route r and r' are two feasible routes for train t, i.e. their incoming and outgoing tracks coincide with the origin and destination of train t in the network. Route r consists of partial routes r_1, r_2, r_3

and r_4, route r' consists of partial routes r'_1, r'_2, r'_3 and r'_4 and r_1 coincides with r'_1. Without using the division in partial routes, both $x_{t,r}$ and $x_{t,r'}$ must be taken into account in the constraints (4.9) for the nodes on $r_1(= r'_1)$. When using this division in partial routes, only x_{t,r_1} must be taken into account in these constraints. Since the number of routes that share a partial route can be very high, this can lead to a huge reduction in the number of terms in the constraints (4.9). This made it computationally possible to include all the routes in the routing model within a reasonable amount of time.

Table 4.9: The number of partial routes in each part of the network in two opposite directions (South to North and North to South).

Network parts			South to North	North to South
Border (South)	-	Brussels-South	1 282	1 824
Brussels-South	-	Brussels-Central	188	189
Brussels-Central	-	Brussels-North	27	28
Brussels-North	-	Border (North)	913	802

Nevertheless, the consecutive setting of partial routes for each train now needs to be included into the routing model. Including this decision into the model, however, not only reduced the model construction time, but also significantly reduced the computation time up to one-hundredth. Constraints (4.8) are replaced by constraints (4.36)-(4.38):

$$\sum_{r \in \mathcal{R}_{\vec{t},\mathrm{dep},s^{\ell_t}}} x_{t,r} = 1 \qquad \forall t \in T \tag{4.36}$$

$$\sum_{r \in \mathcal{R}_{\vec{t},\mathrm{arr},e^{\ell_t}}} x_{t,r} = 1 \qquad \forall t \in T \tag{4.37}$$

$$\sum_{r \in \mathcal{R}_{\vec{t},\mathrm{arr},w}} x_{t,r} = \sum_{r' \in \mathcal{R}_{\vec{t},\mathrm{dep},w}} x_{t,r'} \qquad \forall t \in T, \forall w \in W \setminus \{s^{\ell_t}, e^{\ell_t}\}, \tag{4.38}$$

where s^{ℓ_t} and e^{ℓ_t} are respectively the first and the last node (in the considered network) of the line operated by train t, $\mathcal{R}_{\vec{t},\mathrm{dep},w}$ ($\mathcal{R}_{\vec{t},\mathrm{arr},w}$) contains all partial routes that depart from (arrive in) node w in the direction of train t, indicated with $\vec{t}$. These constraints are flow conservation constraints for each train. They assure that (i) there is exactly one partial route assigned starting at the border node of the incoming line of the train (in the correct direction), constraints (4.36), (ii) there is exactly one partial route assigned arriving in the border node of the outgoing line of the train (in the correct direction), constraints (4.37), (iii) in the intermediate parts of the network, partial routes are assigned such that one feasible, i.e. connected, route is built for each train, constraints (4.38). If a train enters the network in platform area $\mathcal{P}$ (and not in a border node of the network) and the platform is not decided upon beforehand, then for train t we include as many binary variables as there are platforms p in that platform area: $\text{bin}_{p,t} \in \{0,1\}, \forall p \in \mathcal{P}$. We replace constraint (4.36) of train t by the following constraints:

$$\sum_{p \in \mathcal{P}} \text{bin}_{p,t} = 1 \tag{4.39}$$

$$\sum_{r \in \mathcal{R}_{\vec{t},\text{dep},p}} x_{t,r} = \text{bin}_{p,t} \quad \forall p \in \mathcal{P}. \tag{4.40}$$

So the train has to depart from exactly one platform in platform area $\mathcal{P}$, constraint (4.39), and has to be assigned to exactly one route that starts from that platform (and not to one of the routes starting from the other platforms), constraints (4.40). For trains which terminate their route in a platform area without a platform being fixed beforehand, analogous constraints are included.

4.4.2 Constraints to speed up the timetabling model

Based on the output of the routing model and the calculated blocking times, the total time that each node will be blocked in one period of the cyclic timetable can be calculated before the timetable is actually constructed. This information is used to put an upper bound on the minimum buffer time in the timetabling model: the minimum buffer time will always be smaller than the time that a node is not blocked divided by the number of trains that pass in that node.

Secondly, note that two timetables that only differ in the fixed amount of time that is added to each arrival and departure time, are in essence the same. We remove this symmetry from the model by fixing the reservation time of the first section of the train with the smallest index, to zero.

4.4.3 Splitting, coupling and re-usage of trains in routing and timetabling models

In case it is planned to split a train into two parts in a platform area, the first part is scheduled as the original train and the second part is scheduled as a *new* train that starts its route from the platform where the splitting occurs. Each of these two trains has another outgoing line from then on. In case of coupling, the first train arriving at the platform is scheduled to finish its trajectory at the coupling platform and the second train arriving at the coupling platform continues as the coupled train from then on. In case a train is planned to be re-used on a next line after it finishes a first line, both lines use the same physical train. However, we schedule the second trip of this train as a *new* train. This new train starts its route from the platform where the original train ended its route. For the splitting of trains and for the re-usage of a train, the new train needs to start from the same platform as where the original train is split, in case of splitting, or where the original train ended its first trip, in case of a re-usage. Therefore we add the next constraints to the routing model for every splitting and re-usage:

$$\mathrm{bin}_{p,t'} = \sum_{r \in \mathcal{R}_{\vec{t},\mathrm{arr},p}} x_{t,r} \qquad \forall p \in \mathcal{P}, \tag{4.41}$$

where t is the original train and t' is the new train after splitting or re-usage and $\mathcal{P}$ is the platform area where the splitting or re-usage occurs. We use the $\mathrm{bin}_{p,t'}$ variable, because it is not known before the optimization on which platform the new train will start its trip through the network. For the coupling of trains the two trains have to arrive on the same platform. This can also be modeled with constraints similar to constraints (4.41). Corresponding to the data of Infrabel, the time that is scheduled between the release times of the platform of both trains that take part in a splitting is three minutes. In the timetabling model, this can be caught in the following constraint:

$$\mathrm{rel}^{\mathrm{aux}}_{t',p} = \mathrm{rel}^{\mathrm{aux}}_{t,p} + \mathrm{relrel}^{\mathrm{int}}_{t,t',p} + 3, \tag{4.42}$$

where p is the platform in $\mathcal{P}$ where the splitting takes place, which is already assigned by the routing model, t is the original train in the splitting and t' is the new train resulting from the splitting, $\mathrm{relrel}^{\mathrm{int}}_{t,t',p}$ is an integer variable to account for the periodicity. Note that we take the difference between two release times, and not the difference between a reservation time and a release time as in constraints (4.19)-(4.22), such that the $\mathrm{resrel}^{\mathrm{int}}$ variables cannot plainly be used here. That is why we introduced here the $\mathrm{relrel}^{\mathrm{int}}$ variables. It is easy to relate the $\mathrm{resrel}^{\mathrm{int}}$ variables to the $\mathrm{relrel}^{\mathrm{int}}$ variables, but this is not necessary to build a correct model. There is no prescribed time interval between the arrival and departure of the two trains that take part in a re-usage. In order to assure that passengers on the first train do not have to wait too long before the coupling takes place and the coupled train actually leaves the platform, constraint (4.42) can also be used for the two trains involved in the coupling, but with $\mathrm{rel}_{t,p}$ replaced by $\mathrm{res}_{t,p}$ where t is the first train arriving and t' is the second train arriving. Furthermore, the 'three' in constraint (4.42) must be replaced by an appropriate value. Since there are no trains coupling in our case study, we have no data on the coupling time. To prevent that other trains are planned on the platform while the train that will be re-used or coupled has already arrived on the platform, but has not yet left the platform, we include the following constraints in the timetabling model:

$$0 \leq (\mathrm{res}^{\mathrm{aux}}_{t'',p} - \mathrm{rel}^{\mathrm{aux}}_{t,p} + P\,\mathrm{resrel}^{\mathrm{int}}_{t,t'',p}) - (\mathrm{res}^{\mathrm{aux}}_{t',p} - \mathrm{rel}^{\mathrm{aux}}_{t,p} + P\,\mathrm{resrel}^{\mathrm{int}}_{t,t',p}) \leq P, \tag{4.43}$$

where t is the first train in the re-usage or coupling, t' is the second train in the re-usage or coupling and t'' is a third train that is planned to use the same platform on its route. These constraints assure that the reservation time of the second train of the re-usage is closer to the release time of the first train of this re-usage than the reservation time of any other train that makes use of this platform. Note that this constraint is only valid if the index of train t is smaller than that of train t' and t'', since the value and the sign of $\mathrm{resrel}^{\mathrm{int}}_{t,t',p}$ ($\mathrm{resrel}^{\mathrm{int}}_{t,t'',p}$) depends on the order of the indices of train t and t' (t and t''). These constraints are changed accordingly for the other cases, i.e. index of train t bigger than index of train t' but smaller than that of train t'', etc.

4.4.4 Passenger transfers in routing and timetabling models

In case many passengers travel from an origin to a destination for which no direct line is provided in the line plan, a good service provides a smooth transfer from the feeder train to the connection train. In order to facilitate passenger transfers, the feeder and the connecting train could be planned on specific platforms in the routing model, for example adjacent platforms. Let p be a platform in the platform area where the transfer is planned. Define $\mathcal{V}_p$ as the set of platforms for the connecting train which assure a fluent transfer in case the feeder train would arrive on platform p. This fluent transfer is then assured by adding the following constraint for each p in the platform area where the transfer may take place

$$\sum_{r \in \mathcal{R}_{\vec{t},\mathrm{arr},p}} x_{t,r} = \sum_{r' \in \bigcup_{p' \in \mathcal{V}_p} \mathcal{R}_{\vec{t'},\mathrm{dep},p'}} x_{t',r'}, \tag{4.44}$$

where t is the feeder train and t' the connecting train. Furthermore, the transfer time could be bounded by adding the following constraints to the timetabling model:

$$L \leq \mathrm{rel}^{aux}_{t',p'} - \mathrm{res}^{aux}_{t,p} + P\,\mathrm{relres}^{\mathrm{int}}_{t,t',p} \leq U, \tag{4.45}$$

where t and t' are once again the feeder and connecting train, p and p' the platforms assigned in the routing model, $\mathrm{relres}^{\mathrm{int}}_{t,t',p}$ is an integer variable to account for the periodicity and L and U are appropriate bounds for the transfer time. Note that we now take the difference between a release time and a reservation time, and not the difference between a reservation time and a release time as in constraints (4.19)-(4.22), such that again the $\text{resrel}^{\text{int}}$ variables cannot plainly be used here. To this end we introduced the $\text{relres}^{\text{int}}$ variables. Also here it is easy to relate the $\text{resrel}^{\text{int}}$ variables to the $\text{relres}^{\text{int}}$ variables, but this is not absolutely necessary to build a correct model. In these equations, the release variable, rel^{aux}, and the reservation variable, res^{aux}, could also easily be replaced by variables to indicate the arrival and departure times of the trains on the platform. Constraints must then be included that specify the relation between these arrival and departure variables and the rel^{aux} and res^{aux} variables and the appropriate values for L and U must be slightly altered. Unfortunately, there is no data available on passenger transfers in Brussels, so we did not include these constraints when applying our model to the case study.

4.5 Results of the routing and the timetabling model

In this section, we discuss the performance of the developed routing model and timetabling model by applying it to the case study of Brussels. First, the results of the routing model are presented. Then, the delay scenario used in the simulation is specified. Thereafter the timetabling and simulation results are presented.

4.5.1 Routing results

First the routing model (4.5)-(4.13) is tested with different numbers of trains. Secondly, the routing model is tested with alternative objective functions for (4.1) in order to justify the chosen objective function.

Table 4.10 shows the results of the routing model for different amounts of trains (with half of the trains going from Brussels-North to Brussels-South and the other half in the opposite direction). The calculations are done with CPLEX 12.6 on an Intel Core i7-5600U CPU @ 2.60 GHz. The first optimal solution that is found by CPLEX is used to generate the results in Table 4.10. Note that there are trivial alternative optimal solutions in case trains share the same inbound and outbound tracks. For Brussels railway bottleneck there also exists non-trivial alternative optimal solutions, but these are ignored. For the 85 trains case, the results of the first found optimal routing are compared to the results of the reference routing plan from the Belgian railway infrastructure manager Infrabel and from Dewilde (2014b). The values from the latter two are presented in the first two lines of the table. The first column indicates whether constraints (4.7) or (4.4) are used in the routing model. This is not applicable for the reference routing plan of Dewilde (2014b) and Infrabel. The second column contains the number of trains that are routed through the network. The third column denotes the time to construct the routing model. The computation time needed by CPLEX to construct the routing model increases rapidly with the amount of trains, but is still acceptable for the 85 trains case. This is because the number of terms in constraints (4.9) increases with the number of potential routes for these trains. The computation time to solve the routing model, in the fourth column, also increases with the amount of trains, but remains very short. The fifth column shows the maximum node usage. This maximum equals 16 for the 85 trains case and can be explained by the line plan, since 16 trains arrive via the same boundary point in the considered network. The sum of the squared node usages in the sixth column shows an improvement of 13.13% compared to the reference routing plan of Infrabel. The last line presents the results of the routing model with constraints (4.4) instead of constraints (4.7). We see that the model construction time of both alternatives is about the same, but that the computation time to solve the model differs significantly. While our model solves in some seconds, the model with constraints (4.4) has not even converged to an optimal solution within 12 hours.

Table 4.10: Results from the routing model.

Constraints	# Trains	Constr time	Comp time	max g_w	$\sum_{w \in W} g_w^2$	Impro
-	85	Infrabel		16	27 565	-
-	85	Dewilde et al.		16	27 410	0.56%
(4.7)	20	44s	8s	6	1 820	-
	40	232s	14s	8	6 246	-
	60	607s	17s	12	13 397	-
	85	1 403s	26s	16	23 945	13.13%
(4.4)	85	1 377s	>12h	16	23 978	13.01%

Now, different alternative objective functions for (4.1) are tested in order to justify the chosen objective function. Intuitively, one can expect that our routing model does produce short train travel times. This is because the running times of the trains are proportional to the length of the sections on their route. These section lengths are in turn positively correlated to the number of nodes in the section. Thus minimizing the node usages should have a positive effect on the total train travel times. The alternative objective functions that are considered, are:

a) Minimizing the maximum node occupation time.

b) Minimizing the sum of the node occupation times over all nodes.

c) A combination of the previous two criteria (a and b) with a high weight for the first criterion.

d) Minimizing the sum of all train travel times while restricting the node occupation time to at most the period length of the cyclic timetable.

e) Minimizing the sum of the travel times over all trains without restriction on the node occupation time. This can produce a routing plan where a node will be occupied for longer than the period length of the timetable. In this case the routing plan is infeasible for timetabling. However, this is interesting to find out what the real shortest travel times are for every train.

Note that neither of these alternative objective functions, nor our own objective function assures a conflict-free timetable. Table 4.11 presents the results when using these alternative objective functions. Most of the performance indicators in this table are based on the alternative objective criteria. The other performance indicators are the number of non-used platforms in Brussels-South and the numbers of trains per platform in Brussels-South.

Our own objective function leads to a 2.3% higher total travel time of the trains than the overall minimum value (e). Our node occupation time is higher than the lowest value, 5258.7 compared to 4749.0 (c). However, we observe that for objective d) and e), where the travel time of the trains is minimized, the total node occupation time is much higher than for the alternative objective functions. More importantly, from this table, we observe that the maximum node usage is much lower for our original objective function compared to the other objective functions, as was the case for the sum of the squared node usages. We also achieve a more equal spreading of the trains over the platforms, since the maximum number of trains at one platform, is the lowest. Furthermore, we did some extra experiments to construct a corresponding timetable for case b) and we observed that even for this case, where the maximum node usage is the second lowest (25), no feasible timetable is found within 12 hours of optimization. If a certain node is used 25 times (or more) it just becomes very difficult to construct a feasible timetable. So due to the high node usage, none of the other routing plans is suitable for timetabling. This comparison clearly favors and justifies the choice for our original objective function.

Table 4.11: Performance results of different objective functions. Here avg = average, min = minutes, / = per, max = maximum, pl = platform, sq = squared.

Obj.	Total train travel time (min)	Avg train travel time (/train, min)	Total node usage	Sum of sq node usage	Avg node usage (/node)	Max node usage (/node)
(4.1)	2 106.0	24.8	2 833	23 945	5.9	16
a)	2 086.5	24.5	2 760	33 128	5.7	30
b)	2 111.5	24.8	3 222	31 442	6.7	25
c)	2 088.5	24.6	2 771	30 973	5.8	28
d)	2 057.8	24.2	3 953	58 983	8.2	41
e)	2 049.8	24.1	3 809	63 427	7.9	47

Obj.	Total node usage time (min)	Avg node usage time (/node, min)	Max node usage time (/node, min)	Non-used # pl in Br-South	Max # trains/pl
(4.1)	5 258.7	10.9	46.4	1	15
a)	4 811.6	10.0	59.6	10	18
b)	5 687.0	11.8	46.4	2	15
c)	4 749.0	9.9	46.4	7	16
d)	6 974.2	14.5	59.5	3	22
e)	6 629.6	13.8	119.6	7	40

4.5.2 Delay scenario

The delay scenario that is considered in this chapter is one in which approximately 50% of the trains are delayed according to an exponential distribution with an average in the interval [0 min, 8 min]. The approximately 50% of the trains and the averages of the exponential distributions are chosen randomly, but they are fixed in advance. These approximately 50% of the trains and the averages are not linked or correlated to each other. The averages are different for each train. The remainder of the trains are supposed to be on schedule when entering the network.

4.5.3 Timetabling and simulation results

In Table 4.12, we show the robustness results of the solutions of the routing model and the timetabling model. The routing plan used as input for the construction of the timetable is the (first) optimal routing plan produced by our routing model.

This routing plan is described on line 6 of Table 4.10. We experimented with including different amounts of supplements in the timetable: more specifically 0%, 25%, 41%, 50% of the dwell time was added as a supplement. Note that without supplements, a delay can never be absorbed. Including 41% of dwell time supplements gives us a total amount of supplements as close as possible to the timetable of Infrabel and Dewilde (2014b), to which our timetables will be compared. The amount of supplements is not exactly equal due to a time discretization of 0.1 min for these supplements.

We calculated timetables 1-2-3-4 by using a compute node Xeon E5-2680v2, IB-QDR, 20 cores with 16 GB memory for three hours and for timetable 5 with 64 GB memory for 24 hours. Apart from the difference in computation time and available memory between timetables 1-2-3-4 on the one hand and timetable 5 on the other hand, all five timetables are constructed starting from the same routing plan, but differ in the amount of dwell time supplements.

Table 4.12: Simulation results.

Timetable	Supplement (min)		Minimum buffer time (s)
Infrabel	203		-
Dewilde et al.	203		-
1	0	(0%)	5
2	118	(25%)	4
3	206	(41%)	1
4	226	(50%)	0
5	206	(41%)	17

Timetable	Passenger robustness ($\cdot 10^6$ min)	Impro (%)	Nominal travel time ($\cdot 10^6$ min)	Delay out (%)	Deadlocks
Infrabel	2.89	-	1.35	60.19	1 687
Dewilde et al.	2.61	9.43	1.35	49.17	5 174
1	3.33	-15.30	1.34	80.81	47 361
2	2.73	5.31	1.34	59.65	27 229
3	2.56	11.42	1.34	42.99	6 796
4	2.78	3.81	1.34	46.01	49 056
5	2.63	8.83	1.34	41.83	4 696

We see an improvement in passenger robustness (column *'Passenger robustness'*) of 11.42% for timetable 3 (column *'Impro'*) compared to the reference timetable of Infrabel. This is also better than the 9.43% improvement of the timetable found in (Dewilde, 2014b). The column *'Supplement'* reports the amount of dwell time supplements included in the timetables. The more dwell time supplement is added, the less trains leave the network with a delay (column *'Delay out'*) and the smaller the best found minimum buffer time after three hours of optimization (column *'Minimum buffer time'*). Note that these minimum buffer times are very short, but that the safety headways between the trains, defined by the signals, are respected and that these buffer times should be considered as additional time that can be used to avoid the propagation of delays. We see that the passenger robustness is very bad if no supplements are included in the timetable. The passenger robustness improves if more supplements are included, but deteriorates again if 50% or more dwell time supplements are included. If the passenger robustness does not improve any longer by adding more supplements, an equilibrium between supplements and buffer times is achieved.

The column *'Nominal travel time'* shows that our routing model constructed a routing plan with a shorter nominal travel time than that of the routing plan of Infrabel and Dewilde (2014b). This nominal travel time is the planned travel time without weights distinguishing between necessary running and dwell times, supplements and transfer times. For timetable 5, we see that the best minimum buffer time found is much better and that the number of trains which leave the network with a delay improves, while the passenger robustness deteriorates. This can be explained by the fact that the optimization models do not take passenger numbers into account while planning the trains. The impact of a delay on the passenger robustness increases proportionally with the number of passengers on the delayed train.

The last column presents the number of deadlocks reported by the simulation module. This is the number of simulation runs in which two trains are in each other's next section such that neither of them is able to proceed any further on their assigned route. This is called a deadlock situation and then the simulation needs to start over again. Only successful runs are counted in the 10 000 simulation runs. Actually, this is a drawback of the simulation module. The problem is that the simulation cannot solve or prevent deadlocks, it can only report them. The simulation tool cannot predict or handle this situation, since it only contains straightforward conflict resolution techniques. It is difficult to interpret the amount of deadlocks, since the simulation gives no indication on how much (or how little) extra delay would be necessary to avoid them or whether it would be possible to reroute the trains in real time. Furthermore, (i) it can be one or a small number of train pairs that cause a deadlock only in case of larger delays or (ii) it can be a high number of train pairs that already cause a deadlock in case of small delays or (iii) it can be any combination. So a high or low number of 'deadlock runs' in the simulation gives no indication about the quality of a given timetable. This is shown and discussed in Appendix A.1. The conclusion there is that although the actual value of the passenger robustness changes when using different strategies to handle deadlocks, the trend in improvement compared to a reference timetable remains.

These results show that we can construct a routing plan and timetable from scratch which performs better than the reference routing plan and timetable from Infrabel and the best found routing plan and timetable for Brussels from literature, constructed by Dewilde (2014b).

4.6 Extended approach

In this section, we elaborate on an extension of the routing and timetabling model described above. Table 4.12 showed the results when we experiment with different amounts of supplements to include into the timetable. We observed that they largely affect the passenger robustness. The goal of this extended approach is to find the ideal amount and location of supplements in a structured way, in order to avoid the use of random experiments.

An overview of the algorithm is presented in Figure 4.7. The extended approach works iteratively and consists of four pillars. These four pillars are performed consecutively in each iteration. First the input and the output are described, then the four pillars are explained and how they interact with each other.

4.6.1 Input

The input coincides to the input for the routing and the timetabling model described in Section 4.2.1. The only difference is that now the passenger numbers on the lines are considered to be known during the optimization, including the changes in passenger numbers in the different platform areas.

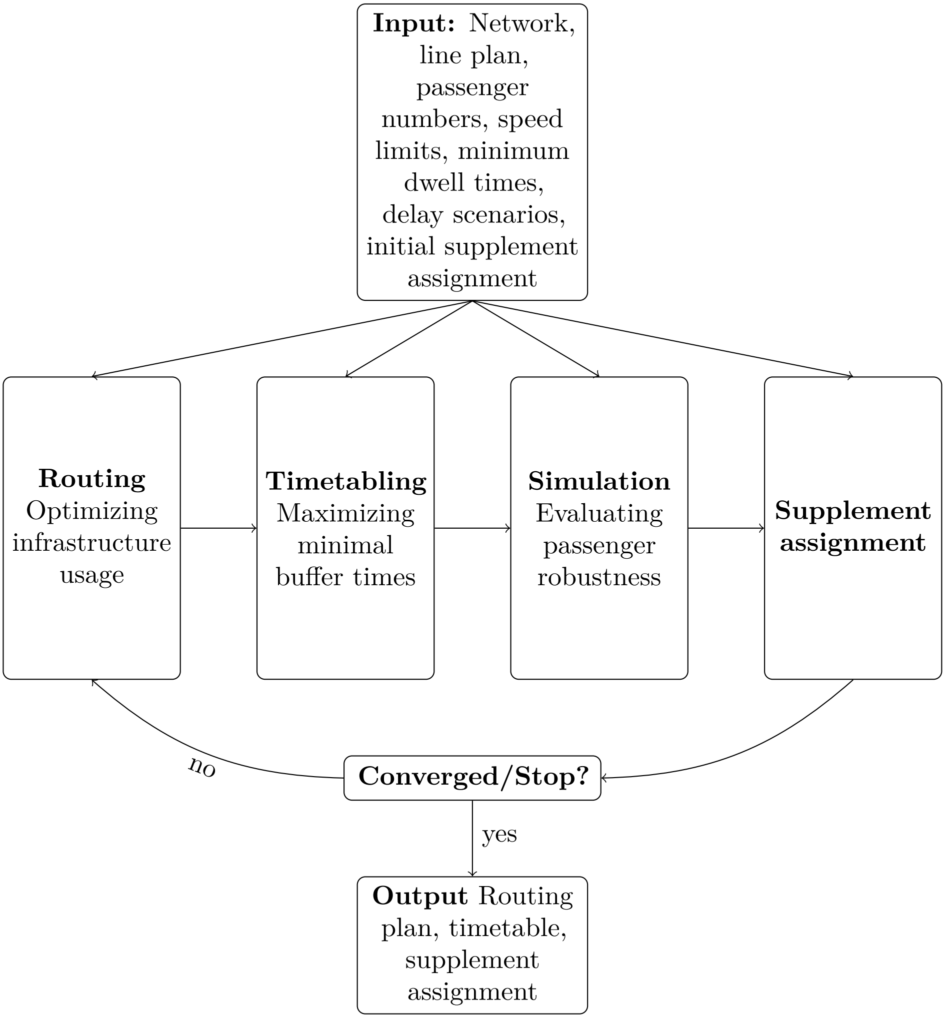


Figure 4.7: Extended approach: algorithm lay-out.

4.6.2 Output

The output consists of the best performing routing plan and timetable, i.e. with the smallest total passenger travel time in practice. An additional output is the final amount and distribution of the supplements.

4.6.3 Algorithm

The four pillars are executed in each iteration until a stop criterion is satisfied. The input that changes in each iteration is the assignment of dwell and running time supplements for each train in each platform area and grid zone. In the first pillar, trains are assigned to routes. Unless the total node usage time for at least one node is not feasible anymore, the routing plan will remain the same over the iterations. The total node usage time can increase and become infeasible due to the new supplement assignment in the current iteration. In that case, a new routing plan is constructed for which the total node usage time does not exceed the threshold value for either of the nodes. In the second pillar, a timetable is constructed. The timetable will change in each iteration since it is directly affected by the new supplement assignment. In the third pillar, the passenger robustness of the constructed timetable and routing plan for the current iteration are computed. This is done by simulation. Since our idea is to use probability distributions of historic delays, the same delay scenario is used during the entire algorithm. In the fourth pillar, a new supplement assignment for the next iteration is made based on the non-used supplements and the propagated delays. After executing the fourth pillar, the stop criteria are checked. We now elaborate on the four pillars and the stop criteria.

Routing

The first pillar is based on the routing model explained in Section 4.2.3. This routing model is extended as follows. Once the trains are assigned to routes, the total node usage time for each node is calculated. This is used to check whether one of these total node usage times exceeds the available time. If so, an extra constraint is included in the model for each of the critical node usage times and the model is solved again. So, suppose that the total node usage time of node $w \in W$ exceeds the available time, for example the period length P or perhaps more realistically, the value that may not be exceeded to stay below the maximum allowed occupation rate. Then, the following constraint for node w is introduced into the model, where we assume that P is the available amount of time:

$$\sum_{t \in T, r \in R_t} q_{t,r,w} x_{t,r} \leq P, \tag{4.46}$$

where $q_{t,r,w}$ is the time that train t blocks node w when using route r and $x_{t,r}$ is as defined in Section 4.2.3. After executing the model again, the total node usage times are checked and extra constraints are constructed if necessary. This process is repeated until the final routing plan optimally spreads the trains over the available infrastructure without exceeding the available time of any node. Even when new additional constraints have to be added several times, the computation time of the routing model remains small compared to the construction of a timetable. Since the number of nodes in a network and the number of terms in constraints (4.46) may be high, it is advantageous that these constraints only have to be constructed if necessary.

Timetabling

The second pillar is based on the timetabling model in Section 4.2.4. This timetabling model is extended by altering the objective function such that passenger numbers can be taken into account. We here explain how this objective function is altered.

The objective function (4.14) is twofold. The first part maximizes the overall minimum buffer time. This part assures a minimum amount of buffer time precursory to any train. This first part remains the same. The second part maximizes the sum of the minimum buffer times between train pairs. Since the main goal is to minimize the passenger travel time in practice, providing sufficient buffer times precursory to highly crowded trains is even more important than further increasing the minimum buffer time precursory to any train. The passengers that are affected by an insufficient buffer time between two trains are the passengers on the second train, since a delay of the first train larger than the buffer time propagates to the second train. This is illustrated in Figure 4.8. That is why weights are now included in the second part of (4.14) to account for the number of passengers that may be affected in case the buffer time is insufficient to avoid delay propagation. So instead of treating all trains equally, trains that transport a high number of passengers are prioritized.

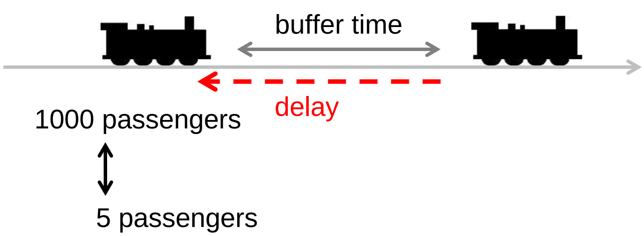


Figure 4.8: An insufficient buffer time between two trains may affect the travel time of the passengers on the second train.

The weights have to satisfy the next two conditions. First, the more passengers that can be affected by delay propagation, the more impact this buffer time must have on the objective function. Secondly, the smaller the buffer time, the more impact this buffer must have on the objective function. The second condition is fulfilled by the original objective function (4.14) without weights. For the first condition, a first intuition could be to use the number of passengers that are affected if the buffer time is insufficient, directly as weight for this buffer time. However, this has not the desired effect. This is due to the fact that the second part of the objective function then maximizes the sum of minimum weighted buffer times between train pairs. The more affected passengers, the higher the weight, the less probable it becomes that the weighted buffer time is also the actual minimum weighted buffer time between these two trains and vice versa. By contrast, using one over the number of passengers that is affected works well. The more passengers may be affected the smaller the weight and thus the larger the impact on the objective function, which maximizes the minimum weighted buffer times. Nonetheless, to avoid fractional coefficients in the model, the buffer times are split into J groups based on the number of passengers that may be affected. This is done by splitting the set from zero up to the maximum number of passengers that may be affected into J disjoint, but consecutive intervals. Group 1 contains the buffer times that may affect the least number of passengers and group J contains the buffer times that may affect the highest number of passengers. Let C be the lowest common multiple of $\{1, \cdots, J\}$ and let the weight for a buffer time B be $C/\text{group}(B)$, where $\text{group}(B)$ is an integer referring to the number of the group of buffer times in which buffer time B is ranked. These weights are integers by construction and satisfy the two conditions. In case the J groups are intervals of equal length, the weighted buffer time decreases linearly with the group number, i.e. the number of passengers that are affected. The weighted buffer times also increase linearly with the magnitude of the buffer time itself. These relations are visualized in Figure 4.9.

The second part of the objective function (4.14) of the timetabling model then becomes:

$$\max \quad \min\{\lambda_{t_i,t_j,w}\,\text{buf}_{t_i,t_j,w} | w \in W, (t_i,t_j) \in T_w \times T_w : t_i \neq t_j\}, \qquad (4.47)$$

where $\lambda_{t_i,t_j,w}$ is the weight of the buffer time between trains t_i and t_j in node w. In Table 4.13 an example is provided, where the maximum amount of affected passengers is limited by 60. The number of affected passengers are assembled per ten in six groups. So here J equals six. These groups are numbered from one to six, from a smaller amount of affected passengers to more affected passengers. The smallest common multiple C of $\{1,2,3,4,5,6\}$ is 60. Suppose that a buffer time B may affect 15 passengers, then it is categorized in group 2, so $\text{group}(B) = 2$. The weight for B then equals $C/\text{group}(B) = 60/2 = 30$.

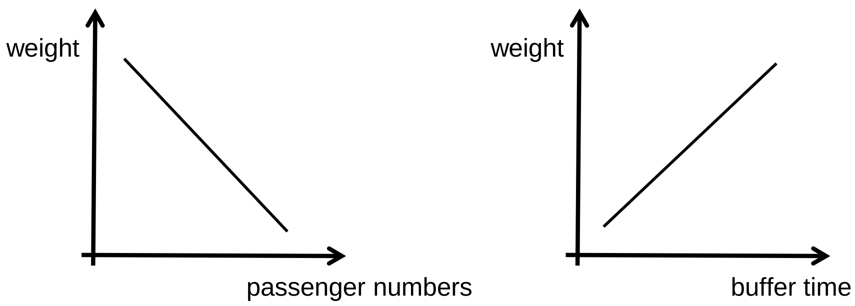


Figure 4.9: The weights decrease linearly with the group number, i.e. the number of passengers that are affected, and they increase linearly with the magnitude of the buffer time itself.

Table 4.13: The number of affected passengers are assembled per ten in six groups. The smallest common multiple of $\{1, 2, 3, 4, 5, 6\}$ is 60.

Possibly affected passengers	$[0, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$	$[50, 60)$
Group	1	2	3	4	5	6
Weight λ	60	30	20	15	12	10

Simulation

The third pillar consists of simulating the routing plan and timetable constructed in the first two pillars. This is done with the simulation tool of Dewilde (2014b), which is described in detail in Section 3.5. The delay scenario coincides with the delay scenario used earlier in this chapter to test the performance of a single routing plan and timetable constructed by the routing and the timetabling model. This delay scenario is described in Section 4.5.2.

Supplement assignment

In the fourth pillar, a new supplement assignment is determined. To assign new supplements, it is useful to keep track of several criteria during the simulation: the total knock-on delay per train, the total knock-on delay that is incurred per signal, i.e. the total amount of time that trains have to wait at this signal when being halted at this signal because their next section is still blocked, the total knock-on delay that is incurred per node, i.e. the total amount of time that trains have to wait in a section containing this node when being halted at the next signal because their next section is still blocked, the total delay incurred per train per node, i.e. the amount of time that a certain train has to wait in the section that contains that node when being stopped by its next signal and the total amount of non-used supplements. These values are calculated for each of the 10 000 simulation runs and the average for each criterion is reported. Furthermore, the distribution of the delays incurred per train per node over the 10 000 simulation runs and the distribution of used supplements per train per platform area and grid zone over the 10 000 simulation runs are determined.

Some or all of these values can be used in various combinations to determine a better assignment of the supplements. For example, the knock-on delay per train can be used to estimate the amount of supplement to be assigned to each train. The knock-on delay incurred at a certain node or signal can be used to estimate the amount of supplements to be assigned in the platform area or grid zone of that signal or node.

After some preliminary tests, the following approach to calculate the next running and dwell time supplements was selected. For each train, the switches and platforms on its route are analyzed in chronological order up to the switch or platform where (i) this train propagates (part of) its delay to the next train or (ii) this train is delayed for the first time. Both cases are now described in detail.

(i) In case the train propagates (part of) its delay to the next train, we determine the delay that corresponds to the x-th percentile (e.g. x equals 50, 75 or 95) of all simulation runs. We set the supplement to a value that prohibits the delay to propagate in these $x\%$ of the simulation runs. We add the determined value as a dwell time supplement to the dwell time at the train's first platform area with an actual stop. So the train's dwell time in that platform area in the next iteration is the dwell time of the current iteration plus this new dwell time supplement. The dwell time in the current iteration consists of the minimum dwell time and any dwell time supplements that are added as original input or in one of the previous iterations. If a train arrives on its platform with a delay, a dwell time supplement can make sure that the train leaves the station on time or with a smaller delay. In case the train is not delayed, the train stands still for the minimum dwell time and its (total) dwell time supplement in that platform area. In our case study, we assume that primary delays are only present when entering the network and do not occur while the trains are already inside the network. So a train that propagates a delay to another train, without first getting a knock-on delay from another train, already suffered from this delay when first entering the considered network. Note that assigning a supplement to the train's first platform area does not avoid delay propagation from this train to other trains in case the delay propagation occurs before the first platform area. However, a supplement may also be assigned to the train that suffers from the knock-on delay, such that the knock-on delay can be immediately absorbed. This is explained in (ii). The advantage of the assigned supplement to the first train is that (part of) its initial delay can already be absorbed in the first platform area without increasing the running times in the (first) grid zone.

(ii) In case a train encounters a delay from another train, we determine the delay that corresponds to the x-th percentile of all simulation runs. We set the supplement to the value that compensates the propagated delay in these $x\%$ of the simulation runs. There is no upper bound restriction. Nevertheless, in the actual results there is no huge increase in the amount of supplements. This can be explained, because in each iteration we give at most one supplement per train and this supplement is based on the train's first secondary delay. So at this point, there is no accumulation of secondary delays yet. If the train encounters this first delay in a station area, we add this value as a dwell time supplement to the dwell time in this station area, otherwise, we add this value to the running time of the train's current grid zone. A running time supplement provides extra time for absorbing a train delay in a grid zone between two platform areas, for example when it is forced to stop at a signal because the next section is not free yet. In case the train is not delayed, it lowers its speed below its normal speed, such that it will not arrive too early on its next platform.

If the determined supplement for the x-th percentile of scenarios is zero, independently of case (i) or (ii), check the next switch or platform on the route of the train until a node in which a delay is encountered or propagated and the determined value for the x-th percentile is not zero.

Until now, only extra supplements are added. Furthermore, the higher x the more supplement will be added and the more difficult it will become to find a feasible timetable in a limited amount of time for very dense networks. Therefore, we will also remove unnecessary supplements. We check for each train and each grid zone and platform area that it crosses, which amount of the provided supplement remains unused in $y\%$ of the simulation runs. We remove this supplement. The lower y, the more supplement is removed and the less delays can be absorbed. In case y is chosen equal to 100, even supplements that proved useful in only one simulation run, are not removed. In fact, y has to be chosen such that the scenarios in which a delay could be absorbed are weighted against the scenarios in which the supplement is a superfluous lengthening of the passenger travel time. In the case study, x varies over 50, 75 and 95 and y is chosen equal to 95.

Stop criteria

The iterations over these four pillars stop if one of the following stop criteria is satisfied. First, the approach terminates if the supplement assignment converges, i.e. in two succeeding iterations the supplement assignment is equal. This is the first stop criterion. We choose to execute at most ten iterations, which limits the total duration of the algorithm. This is the second stop criterion. We put a time limit on the timetabling phase, since the time needed for the other phases is small compared to the time needed for the construction of an optimal or even a near optimal timetable. This optimization time depends of course on the network characteristics and on the number of trains. The more complex the network or the more trains, the more difficult it becomes to find a feasible timetable. In case of a dense and complex network, a feasible solution could be far from optimal. Anyway, for our case study, in case a feasible, let alone an optimal timetable, is found within three hours, the algorithm proceeds with the simulation, otherwise, if no feasible solution is found within three hours, the algorithm stops. This is the third stop criterion. If in two successive iterations the passenger robustness is worse than the best found value, no further iterations are done. This is undesired but possible nevertheless, since we do not directly optimize passenger robustness. This is the fourth stop criterion.

4.7 Case study of the extended approach

The same case study for Brussels is used to measure the performance of the extended approach. This case study is described in Section 3.6 and 4.3. However, to generate the first results, only 40 of the 85 trains are taken into account and scheduled. In this scenario 20 trains are driving from Brussels-North to Brussels-South and 20 the other way round. With only 40 trains, the network is not densely occupied, such that trains can be more easily spread. This new setting will have an impact on the results. Later on the results are shown for all 85 trains.

The initial supplement assignment that is (i) used for the feasibility check built in in the routing model and (ii) used to determine the duration parameters of the timetabling model, coincides with common practice, which is 7% (Bešinović et al., 2016a, Caimi, 2009). So each dwell time is lengthened with 7% and the running time (drive time + (if included) time to decrease and increase speed) for each block section not containing a platform is lengthened with 7%.

4.8 Results of the extended approach

Here the results of the extended approach are presented. The delay scenario coincides with the delay scenario described in Section 4.5.2. This section concludes with a discussion of the results.

In Table 4.14 and 4.15 the passenger robustness, presented in the column *'Passenger robustness'*, is the most important performance indicator. The column *'Impro'* presents the percentage improvement in passenger robustness compared to the outcome constructed with the initial supplement assignment for the case with 40 trains. For the 85 trains case, Table 4.15 gives the percentage improvement compared to the reference timetable and routing plan from Infrabel. The column *'Supplements'* presents the amount of supplements present in the timetable and the column *'Non-used supplements'* presents the average amount of supplements that is not used in the 10 000 simulation runs. The column *'Knock-on delays'* shows the average amount of delays that is propagated from one train to the next train over all trains, expressed in train minutes. The next column *'% Delay out'* is the average number of trains that leaves the network with a delay and the column *'% Absorb delay'* gives the average number of trains that absorbs its initial delay before leaving the network. The column *'Deadlocks'* is the number of simulation runs in which two trains are in each other's next section such that they both cannot proceed anymore on their assigned route. This is a drawback of the simulation tool and is explained in Section 4.5.3. The column *'Minimum buffer time'* in Table 4.14 is the minimum buffer time that is achieved for the 40 trains case within a one hour optimization time restriction for the timetabling model, computed with CPLEX 12.6 on an Intel Core i7-5600U CPU @ 2.60 GHz. The column *'Minimum buffer time'* in Table 4.15 is the minimum buffer time that is achieved for the 85 trains case. Here, we use a three hour optimization time restriction for timetabling on a compute node Xeon E5-2680v2, IB-QDR, 20 cores with 16 GB memory.

4.8.1 Results for limited train occupation

Now the results generated for 40 trains applied to the Brussels case study are discussed. The results are presented in Table 4.14. To determine the supplements of the next iteration $x = y = 95$ is used (see Section 4.6.3). The algorithm stops after ten iterations, since two stop criteria are then satisfied. On the one hand, the maximum number of iterations is reached, but on the other hand, the passenger robustness deteriorates in each of the last two iterations. The final outcomes are the routing plan, timetable and supplement assignment of iteration 8, since there the passenger robustness is the best. During the first six iterations the passenger robustness improves with 10.3%. In the seventh iteration the passenger robustness deteriorates, then it improves again in the eighth iteration and thereafter deteriorates again in the next two iterations. The amount of included supplements and non-used supplements both decrease in the first iteration, but then increase. The average percentage of used supplement, equaling 32% ($= \frac{59.8-40.4}{59.8}$) for the initial timetable, lies in between 40% and 45% during the first six iterations and then decreases again. The average knock-on delay decreases first with its lowest value in the fourth iteration. It then increases, but achieves another low value in the eighth iteration. The average percentage of trains that leaves the network with a delay, decreases over all the iterations and the average percentage of trains that absorbs their initial delay increases.

These results imply that the passenger robustness can be highly improved by wisely adding supplements. Remember however, that the network is not densely occupied in this case. Since the minimum buffer time is about the same for all iterations, the improvement can be ascribed to the better amount and assignment of supplements. Also the amount of trains that absorb their initial delay and leave the network with a delay can be highly improved. However, we observe that while the amount of trains that leave the network with a delay keeps improving, this is not the case for the passenger robustness. A possible explanation here could be that we did not take passenger numbers into account while adding supplements. So extra supplements that avoid train delay only have a small effect on the passenger robustness if only a low number of passengers is affected by this delay. Note also that supplements lengthen the travel times of the trains in the network which increases the occupation of the network. If supplements are well placed, then this higher occupation of the network does not increase the probability of knock-on delays in the network. This is because a train can immediately absorb its delay in case a supplement is provided, such that this delay will not be propagated. Knock-on delays are only caused if the delay is larger than the supplement and the consecutive buffer time.

The last row of Table 4.14 presents the simulation results for iteration 8, where we use a computation time of 12 hours instead of only one hour for constructing an optimal timetable. This is indicated as iteration 8*. We see that the minimum buffer time in the timetable improves to 3.1 minutes and that the passenger robustness improves even further (12.1% to 13.2%). Note that all other performance indicators also improve, except for the average percentage of trains that absorbs its initial delay.

Table 4.14: Simulation results for 40 trains, computed in one hour of optimization time with CPLEX 12.6 on an Intel Core i7-5600U CPU @ 2.60 GHz. *: 12 hours of computation time.

Iter	Passenger robusteness ($\cdot 10^6$ min)	Impro (%)	Supplements (min)	Non-used supplements (min)
0	1.16		59.8	40.4
1	1.12	3.4	55.9	33.1
2	1.10	5.2	62.2	35.9
3	1.07	7.6	65.2	37.7
4	1.05	9.6	66.4	37.6
5	1.05	9.7	70.1	39.1
6	1.04	10.3	79.3	46.5
7	1.04	10.2	89.9	54.1
8	**1.02**	**12.1**	**105.4**	**68.3**
9	1.03	11.8	105.4	67.4
10	1.06	8.7	112.3	74.7
8*	1.01	13.2	105.4	68.9

Iter	Minimum buffer time (min)	Knock-on delays (min)	% Delay out (min)	% Absorb delay (trains)	Dead-locks
0	2.7	6.3	29.3	17.0	0
1	2.8	6.5	28.3	18.7	238
2	2.9	5.0	24.4	21.9	0
3	2.8	3.9	21.6	23.9	0
4	2.9	2.9	19.5	25.2	493
5	2.8	4.4	19.7	26.9	771
6	2.8	4.9	18.2	27.8	0
7	2.8	8.3	17.8	29.0	0
8	**2.7**	**3.7**	**13.8**	**31.7**	**170**
9	2.4	6.1	14.4	31.6	258
10	2.6	3.4	13.2	32.5	581
8*	3.1	2.5	13.3	31.2	0

4.8.2 Results for dense train occupation

Table 4.15 shows the simulation results of the extended approach applied to the case where all 85 trains are planned. The performance is compared to the

Table 4.15: Simulation results for 85 trains, the timetables are computed in three hours of optimization time with CPLEX 12.6 on a compute node Xeon E5-2680v2, IB-QDR, 20 cores with 16 GB memory.

Iter	Passenger robustness ($\cdot 10^6$ min)	Impro (%)	Supplements (min)	Non-used supplements (min)
Infrabel	2.89		203	122
Dewilde et al.	2.61	9.4	203	140
Original	2.56	11.4	206	105
0	2.66	8.0	123	61
95% - 1	**2.38**	**17.6**	**177**	**100**
75% - 1	2.84	1.5	133	57
75% - 2	2.63	8.9	170	83
50% - 1	2.57	10.9	118	54
50% - 2	2.69	6.7	127	53
50% - 3	3.05	-	128	51

Iter	Minimum buffer time (min)	Knock-on delays (min)	% Delay out (trains)	% Absorb delays (trains)	Dead-locks
Infrabel	0.0	150	60.2	9.7	1 687
Dewilde et al.	0.0	87	49.2	12.8	5 174
Original	0.0	117	43.0	18.3	6 796
0	0.0	105	55.7	10.1	39 377
95% - 1	**0.0**	**83**	**48.2**	**15.6**	**26 651**
75% - 1	0.0	149	60.9	10.6	17 635
75% - 2	0.0	128	53.9	14.0	22 140
50% - 1	0.0	102	54.5	11.6	17 797
50% - 2	0.0	120	55.1	12.7	40 698
50% - 3	0.0	178	57.0	12.7	31 955

performance of the routing plan and timetable of Dewilde (2014b), presented in the row labeled '*Dewilde et al.*', of Infrabel, presented in the row labeled '*Infrabel*', and the performance of the routing plan and timetable constructed by the routing and timetabling model described in Section 4.2.3 and 4.2.4, presented in the row labeled '*Original*'. The latter routing plan and timetable differ from the routing plan and timetable constructed in the first iteration of the extended approach by the changes made to these models, see Section 4.6.3, and by the amount of supplements that is included. The routing model now includes a feasibility check and the timetabling model takes passenger numbers into account. The parameters of Section 4.6.3 can be interpreted as follows. The higher x, the more simulation runs in which delays early in the network can be avoided, since the amount of supplements matches or exceeds the amount of delays in more simulation runs. The higher y, the less supplements that proved useful in some simulation runs, are removed. As for the case with 40 trains, we first used $x = y = 95$. The resulting passenger robustness of the first iteration was incredibly good, see row *'95% - 1'*. It improved the passenger travel time in practice of Infrabel with over 17%. However, during the second iteration, no feasible timetable could be determined anymore within a three hour computation period, given the added supplements. For $x = 75$ and $y = 95$, the passenger robustness in the first iteration deteriorates, but improves in the second iteration, see row *'75% - 1'* and *'75% - 2'*. The third iteration is however not feasible anymore within three hours of computation time. For $x = 50$ and $y = 95$, the passenger robustness improves in the first iteration, but then deteriorates in the next two iterations, see row *'50% - 1'*, *'50% - 2'* and *'50% - 3'*.

Figure 4.10 shows the blocking times in Brussels-Central station for the reference timetable from Infrabel and Figure 4.11 for the reference timetable from Dewilde (2014b). The labels on the horizontal axis are expressed in minutes. The vertical axis labels the six platforms of Brussels-Central. Each train has its own color and the train number is presented inside the colored rectangle that presents its blocking time. Since each train passes at most once in Brussels-Central per

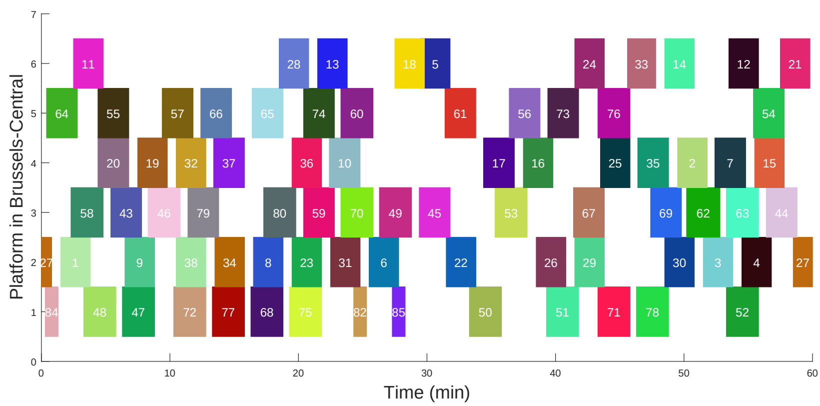


Figure 4.10: The blocking times of the reference timetable of Infrabel for all trains passing Brussels-Central during morning peak hour between 7 a.m. and 8 a.m.

period, at most one blocking time per train is shown.

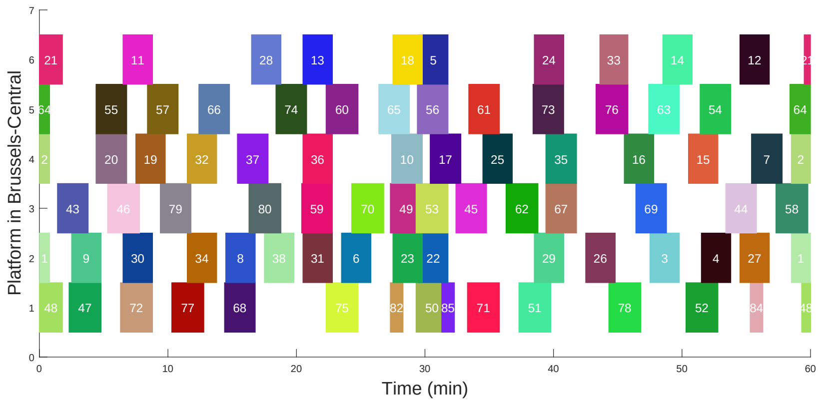


Figure 4.11: The blocking times of the timetable of Dewilde (2014b) for all trains passing Brussels-Central during morning peak hour between 7 a.m. and 8 a.m.

Observe that only one train in the timetable from Dewilde (2014b) has a different platform in Brussels-Central compared to the timetable from Infrabel (train 63). Furthermore, there are several trains that changed order in the timetable from Dewilde (2014b) compared to the timetable from Infrabel, for example train 56 and 61.

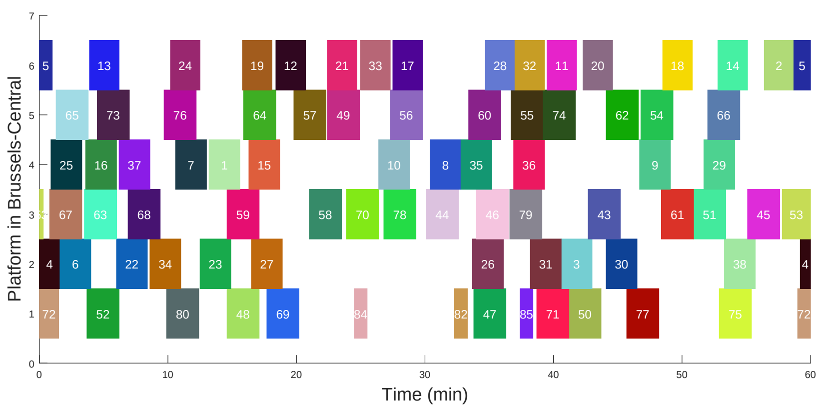


Figure 4.12: The blocking times of the timetable of row *'95% - 1'* for all trains passing Brussels-Central during morning peak hour between 7 a.m. and 8 a.m.

In Figure 4.12, the blocking times in Brussels-Central station are shown for the timetable of row *'95% - 1'* in Table 4.15. The platforms are used 11 up to 15 times. None of the blocking times overlap, so the constructed timetable is conflict-free on the signaling level. Furthermore, we observe that between most blocking times there is a positive buffer time.

At first sight, the timetables from Infrabel and Dewilde (2014b) seem to have a better spreading in Brussels-Central. However, an important remark to make is that these timetables are not entirely cyclic. This cannot clearly be observed from the above diagrams, but it can be noticed when evaluating other nodes. Our timetable is entirely cyclic and it is more passenger robust for the whole station area.

4.8.3 Discussion

For the case with 40 trains, where the occupation rate of the network is only moderate, we observe that the passenger robustness can be highly improved by the extended approach, up to 13.2%. Note that the timetable of the first iteration contains less supplements than the initial timetable and a 3.4% better value for the passenger travel time in practice. This shows that a good implementation of supplements is worthwhile. Furthermore, we see that the minimum buffer time over all iterations is more or less equal, so the improvement in passenger robustness is due to the smart allocation of supplements. Ideas to further improve the supplement assignment are suggested in Section 4.10.

For the case with 85 trains, we also observe that a smart allocation of supplements can highly improve the passenger robustness. However, due to the restriction on the computation time for timetabling, the advantage of a better supplement allocation is counteracted by a bad buffer time assignment. Either the timetable optimization is broken off too early or a feasible timetable is not found within the fixed computation time. There are several options to cope with this problem. A simple solution would be to extend the time limit. Another option is to choose different values for the x and y parameters that determine how much the supplement of the next iteration depends on the simulation results of the current iteration. A third option would be to choose an alternative optimal routing plan as input. A fourth option is to replace the exact timetabling model with a fast heuristic to overcome this problem. This is also suggested and further discussed in Section 4.10.

4.9 Conclusion

In this chapter, we presented a routing model and a timetabling model to construct a conflict-free and passenger robust routing plan and timetable from scratch for complex railway station areas. The focus of these models is on optimally using the infrastructure in the bottleneck area and spreading the trains in time such that the passenger travel time in practice is optimized. We validated these models by constructing a conflict-free routing plan and timetable for Brussels complex railway station area. Furthermore, we compared the performance of this timetable and routing plan with a reference timetable and routing plan from the Belgian railway infrastructure manager Infrabel and with the best timetable from literature for this area constructed by Dewilde (2014b). We also justified to use the node usage as objective function by comparing different objective functions on several performance criteria. The node usage is significantly decreased by 13%. Moreover, simulation of the constructed routing plans and timetables show that our approach outperforms the passenger robustness of the reference routing plan and timetable from Infrabel with up to 11%. This is also 2% better than the best routing plan and timetable from Dewilde (2014b) for this area.

This chapter also presents an iterative approach to construct a railway routing plan and timetable from scratch with a smart allocation of buffer times and supplements. In each iteration, the supplement assignment is refined based on simulation results and the buffer times are optimized in an exact optimization model. The results show that the passenger robustness can be highly improved. However, for very dense and complex networks, a time constraint to construct a well performing timetable can counteract the positive effects of a better supplement assignment. The timetable quality has to be weighed against the computation time. Nevertheless, the iterative approach constructed a timetable and routing plan that improved the passenger robustness of a reference planning from practice with up to 17.6%.

4.10 Future research

An idea to further improve the extended approach is to include during the course of one iteration multiple supplements for one train (at different locations). It might then be necessary to provide an upper bound on the amount of supplements that can be added in each iteration or per train, in order to avoid an overgrowth of supplements. The more supplements, the higher the capacity usage, the more difficult it becomes to construct a conflict-free timetable. A second idea is to make the supplement assignment in the extended approach dependent on the number of passenger travel times that will be affected.

Another idea for future research is to include the assignment of supplements into the timetabling model.With the current objective to maximize the buffer times, the model will generally not include running and dwell time supplements in order to keep the buffer times as long as possible. Supplements will only be added to increase buffer times somewhere further down the network and not necessarily where they are useful in order to absorb delays. Since the timetabling model is already slow now, a fast heuristic is perhaps more appropriate.

A stochastic passenger travel time objective for microscopic timetabling, i.e. the optimization of the passenger travel time in stochastic circumstances (on the microscopic scale), could weigh the advantages of buffer times against (dwell time) supplements in the construction of a timetable. However, constructing an exact model with this objective that can be solved in a reasonable amount of time for microscopic timetabling in complex railway station areas will constitute a big challenge. So developing a fast and well-performing heuristic to construct a passenger robust timetable for dense and complex station areas is a worthwhile research direction. Focusing on the planning in Brussels-Central and the grid zones at both sides of this platform area seems a promising starting position. Trains using the same platform in Brussels-Central could be equally distributed over the length of the cyclic timetable. Thereafter it could be checked whether there are conflicts in the grid zones (and later in other stations). In case of a conflict, either the order of trains could be changed, or depending on the blocking time overlap, trains could be shifted in time. An algorithm based on these principles is constructed for the DSB S-tog network in Copenhagen. For Brussels, however, this is only a first idea, there is no proof of concept yet. Another idea would be to use a decomposition to optimize the exact model. A subset of the constraints contain only train specific variables, however, the larger part of constraints can be expected to be coupling constraints.

4.11 Acknowledgments

This research was supported in part through computational resources provided by the KU Leuven high performance cluster. I thank the Belgian railway infrastructure manager Infrabel for their commitment during the research and for providing data. I also thank Gábor Maróti for the fruitful discussion on the linearization of the routing objective function during the CASPT2015 conference in Rotterdam.

5 Integrating robust timetabling in line plan optimization

This chapter proposes a heuristic algorithm to build a railway line plan from scratch that minimizes passenger travel time and operator cost and for which a feasible and robust timetable exists.[1] A line planning module and a timetabling module work iteratively and interactively. The line planning module creates an initial line plan. The timetabling module evaluates the line plan and identifies a critical line based on the minimum buffer times between train pairs. The line planning module proposes a new line plan in which the time length of the critical line is modified in order to provide more flexibility in the timetabling process. This flexibility is used in the timetabling module in order to improve the robustness of the railway system. The algorithm is validated on the Copenhagen DSB S-tog network, which is a high frequency railway system, where overtaking is not allowed. This network has a rather simple structure, but is constrained by limited shunt capacity. Limited shunt capacity refers in this dissertation to the

[1] This chapter contains the peer-reviewed journal paper:

Sofie Burggraeve, Simon H. Bull, Pieter Vansteenwegen, and Richard M. Lusby. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160, 2017.

and the extended abstract of the TRISTAN IX conference:

Sofie Burggraeve, Simon H. Bull, Richard M. Lusby, and Pieter Vansteenwegen. Train turn restrictions and line plan performance. In *Proceedings of the Ninth Triennial Symposium on Transportation Analysis (TRISTAN IX), Oranjestad, Aruba*, 2016.

Most sentences are cited from the paper and the extended abstract, but in order to keep the text easily readable, cited sentences are not separately indicated.

absence of shunting space in the terminal stations. As a consequence, if a train ends its trip on a platform in its terminal station, no other train can enter this terminal station via the same platform before the first train leaves the station. While the operator and passenger cost remain close to those of the initially and (for these costs) optimally built line plan, the timetable corresponding to the finally developed robust line plan significantly improves the minimum buffer time, and thus the (passenger) robustness, in eight out of ten studied cases.

5.1 Introduction

Traditionally, a railway line plan is constructed before a timetable is made. However, an optimal line plan does not guarantee an optimal or even a feasible timetable (Kaspi and Raviv, 2013). An integrated approach can overcome this problem. Nevertheless, since both line planning and timetabling are separately already very complex problems for large railway networks (Michaelis and Schöbel, 2009, Goerigk et al., 2013), solving the resulting integrated problem up to optimality is not computationally possible in most cases (Schöbel, 2015). We propose a heuristic algorithm that constructs a line plan for which a feasible timetable exists. We call a line plan *timetable-feasible* if there exists a normative macroscopically feasible timetable for that line plan. Moreover, the algorithm strives for passenger robustness. However, aiming for passenger robustness is not explicitly tackled, but implied by focusing on buffer times in order to avoid delay propagation. If delays are less likely to be propagated between trains, fewer passengers will be delayed, which positively affects the total passenger travel time in practice. More specifically, after an initial timetable-feasible line plan is constructed, the passenger robustness is improved by making well-chosen changes in the stopping patterns of the lines while the existence of a feasible timetable remains assured.

This chapter explains the developed approach to build a line plan and timetable from scratch while taking passenger robustness into consideration. The focus is on the interaction between both planning problems. The aim of the approach is to develop a line plan that guarantees a feasible and passenger robust timetable. The approach includes a line planning model that optimizes a weighted sum of the passenger travel time and the operator cost. It also includes a timetabling model that is based on the Periodic Event Scheduling Problem (PESP), to create passenger robust timetables. The integration consists of a smart and targeted interaction between both exact models. A line plan, optimal for a weighted sum of passenger and operator cost, is constructed and iteratively updated until a normative macroscopically feasible and passenger robust timetable is found, while keeping the quality of the line plan high. An advantage of this integrated approach is that it can also be used to improve the robustness of an existing line plan.

The main contributions of the research presented in this chapter are:

- The integration of line planning, timetabling and passenger robustness.
- An approach that builds coordinated line plans and timetables *from scratch.*
- Two insights and proofs on timetable-infeasibility of line plans.
- The inclusion of limited shunt capacity at terminal stations in line plan and timetable optimization.
- Practical conclusions for the DSB S-tog network in Copenhagen based on experimental results.

The context of this research is a high frequency network. The approach is suitable for large networks as long as they have a simple structure, as defined in Chapter 2. Furthermore, the approach is suitable in case trains are forced to turn on their platform in their terminal stations, due to a lack of shunting area.

The proposed integrated approach originates from insights on why some line plans do not allow feasible timetables and why some line plans allow more robust timetables. A first insight is that a line can be infeasible on its own, which we call *line infeasibility.* A second insight is that line combinations can be infeasible due to their frequencies. We call this *frequency combination infeasibility.* In Section 5.2 these insights are explained. The integrated approach itself is presented in detail in Section 5.3 together with the underlying line planning and timetabling module. Thereafter the integrated approach is validated on a case study on the S-tog network in Copenhagen (Denmark). The case study is described in more detail in Section 5.4. In Section 5.5 the results of the case study are presented and examined and the integrated approach is illustrated in an example. This section also shows results on how taking limited shunt capacity into account restricts the potential to construct a passenger robust timetable. The chapter is concluded and ideas for future research are suggested in Section 5.6 and 5.7.

5.2 Timetable-infeasibility

This section explains how limited shunt capacity and certain frequency combinations of lines can lead to timetable-infeasibility of line plans. The integrated approach in Section 5.3 uses these insights to construct line plans that allow normative macroscopically feasible and passenger robust timetables.

5.2.1 Line infeasibility

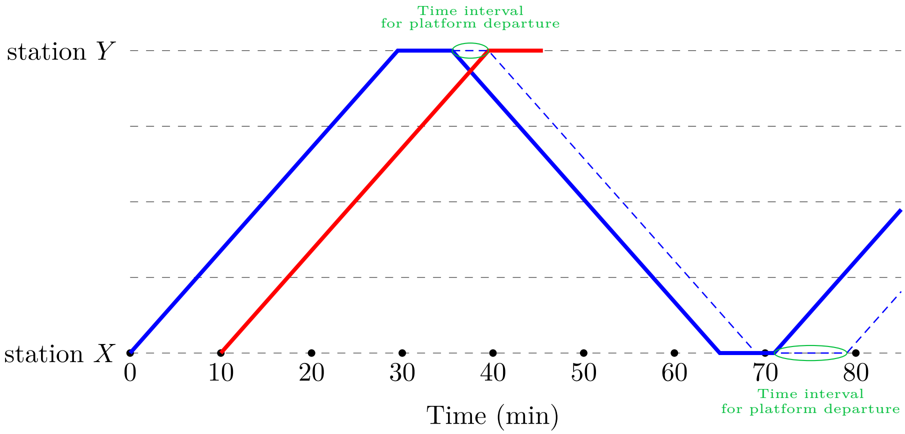


Figure 5.1: A line can be infeasible on its own.

Consider Figure 5.1, showing a single train line operating at a frequency of six times per hour between terminal stations X and Y. Stations X and Y are indicated on the vertical axis. The black dots on the (horizontal) time-axis show the scheduled departures from station X for this line, once every ten minutes. The first two time-distance graphs are illustrated; the first departing from station X at minute zero (solid blue line), and the subsequent train following at minute ten (orange-red line). In this example, the travel time between stations X and Y for the line is 29 minutes. This travel time includes the running times between the stations and the occupation times of the intermediate stations (not in the terminal stations). We assume that the train has to turn on its platform in station X and Y due to restricted shunt capacity and that trains of the same line share the same platform in their terminal stations. The subsequent train that has departed ten minutes later is therefore entering the same station area in Y ten minutes later as well. As a consequence the first train has a well-defined latest departure time of that station area, which is marked as a dashed blue line. The necessary turn time for this line in station Y is seven minutes. Note that the necessary turn time already includes the occupation time of the terminal station Y for the arriving train and the train driving back to station X, which share the same rolling stock. So, this train is arriving in station X again between 65 minutes and 68 minutes after its first departure at minute zero. The necessary turn time in station X is also seven minutes for this line. Thus, the train can leave station X for the next round trip 72 minutes after minute zero at the earliest (minute 65 arrival with seven minutes minimum necessary turn time) and 78 minutes at the latest (68 minute arrival with a maximum of ten minutes for dwelling and turning, assuming that the next train enters the same station area in X ten minutes later). However, no train of this line is planned to leave station X in the interval of 72 to 78 minutes, which can be seen in Figure 5.1: no black dot is situated in the interval surrounded by the green line. Therefore no feasible timetable can be found for this line under the current assumptions and restrictions. We will call this *line infeasibility.*

This insight can be mathematically formulated as: *If there exists no* $k \in \mathbb{Z}^+$ *such that*

$$2\mathcal{T}_l + \mathrm{ntt}_{s_{l,0}} + \mathrm{ntt}_{s_{l,|\mathcal{S}_l|}} \leq \frac{P}{f_l}k \tag{5.1}$$

and

$$\frac{P}{f_l}k \leq 2\mathcal{T}_l + 2\frac{P}{f_l} \tag{5.2}$$

are satisfied, then, in case of restricted shunt capacity in its terminal stations, line l *is infeasible on its own. Here* $\mathcal{S}_l = \{s_{l,0}, \cdots, s_{l,i}, \ldots, s_{l,|\mathcal{S}_l|}\}$ *is the set of all stations on line* l *(independent of an actual stop),* $\mathrm{ntt}_{s_{l,0}}$ *and* $\mathrm{ntt}_{s_{l,|\mathcal{S}_l|}}$ *are respectively the necessary turn time for line* l *in its station of origin* $s_{l,0}$ *and terminal station* $s_{l,|\mathcal{S}_l|}$*,* f_l *is the frequency of line* l*,* P *is the period length of the cyclic timetable and* $\mathcal{T}_l = \sum_{i=0}^{e-1} \mathrm{run}_{l,s_{l,i},s_{l,i+1}} + \sum_{i=1}^{e-1} \mathrm{occ}_{l,s_{l,i}}$ *is the travel time of line* l*, where* $\mathrm{run}_{l,s_{l,i},s_{l,i+1}}$ *consists of the running time between station* $s_{l,i}$ *and* $s_{l,i+1}$*, and* $\mathrm{occ}_{l,s_{l,i}}$ *is the occupation time of station* $s_{l,i}$*. Furthermore, it is assumed that trains of the same line are equally spread over the period and use the same station area in the terminal stations for passenger convenience.*

So in the example above, $\mathcal{T}_l$ is 29 minutes, $\mathrm{ntt}_{s_{l,0}}$ and $\mathrm{ntt}_{s_{l,|\mathcal{S}_l|}}$ are seven minutes, P is 60 minutes and the line frequency f_l is six.

Proof. We define a *train cycle* of line l as (i) the trip from its station of origin to its terminal station including running and dwelling, (ii) the turn movement in its terminal station, (iii) the trip from its terminal station back to its station of origin including running and dwelling and (iv) the turn movement in its station of origin before the train can start a next cycle. The shortest possible duration of a train cycle of line l is the sum of the running and occupation times from the station of origin to the terminal station, $\mathcal{T}_l$, the necessary turn time in its terminal station, $\mathrm{ntt}_{s_{l,|\mathcal{S}_l|}}$, the travel time from the terminal station to the station of origin, $\mathcal{T}_l$ (the travel time is the same in both directions) and the necessary turn time in its station of origin, $\text{ntt}_{s_{l,0}}$. Note that the occupation times of the terminal stations, $\text{occ}_{s_{l,0}}$ and $\text{occ}_{s_{l,|S_l|}}$, are not included in $\mathcal{T}_l$. This shortest possible train cycle length is given in the left hand side (lhs) of formula (5.1). The difference between the longest possible duration and the shortest possible duration is characterized by the time that the train takes for turning in its terminal stations. The time that the train takes for turning in a terminal station is bounded below by the necessary turn time in that station. Furthermore, the train may only stay in the station area until the next train of the same line arrives, which is P/f_l minutes after its own arrival. This P/f_l minutes also includes the occupation time of the arriving and departing train (same rolling stock). This maximal train cycle length is represented in the right hand side (rhs) of formula (5.2). Without loss of generality we can assume that train cycles of line l start at $\{kP/f_l \mid k \in \mathbb{Z}^+\}$. If line l is feasible, then for a train that starts its first cycle at k_0P/f_l for a $k_0 \in \mathbb{Z}^+$, there has to exist a $k \in \mathbb{Z}^+$ for the start of its next cycle such that $kP/f_l \in [k_0P/f_l + (\text{lfs of (5.1)}), k_0P/f_l + (\text{rhs of (5.2)})]$. Note that the latter statement remains true if k_0P/f_l is subtracted from both interval bounds. This proves our mathematical insight by contraposition. As shown in the example, such a k does not always exist. $\square$

5.2.2 Frequency combination infeasibility

Suppose that two lines share a part of the network and that trains of the same line are equally spread in the cyclic timetable. A second insight is that the frequencies of these lines affect the best value (i.e. largest value) of the minimum buffer time between these lines. This best value is tight in case the minimum buffer time between the two lines is maximized. It is straightforward to notice that the higher the frequencies the smaller this best value. But we also make the following claim:

Claim 1. *The best value of the minimum buffer time between a line at a higher frequency and a lower frequency is never greater than between two lines at the higher frequency.*

Example Let $f_l \leq f_{l'}$ be the frequencies of two lines l and l' respectively. If $f_l = f_{l'} = 5$, then on a given infrastructure resource, trains of line l and l' could be planned alternately every six minutes. Without loss of generality, we here assume occupation intervals of length zero, since any larger occupation interval will simply induce smaller buffer times. However, if we assume $f_l = 4$ and $f_{l'} = 5$, then, at any infrastructure resource shared by line l and l' and exactly once in the period of the timetable, there will be two succeeding trains of line l' which are planned between two succeeding trains of l (pigeon hole principle). We will refer to the concerning trains of line l and l' in this event as $t^1_{l,r}, t^2_{l,r}, t^1_{l',r}$ and $t^2_{l',r}$ respectively, where l and l' are the lines concerned, r represents the shared infrastructure resource and the superscript indicates the order of the trains: $t^1_{l,r}$ ($t^1_{l',r}$) proceeds train $t^2_{l,r}$ ($t^2_{l',r}$). In this example, in order to equally spread the trains of the same line, the time between $t^1_{l,r}$ and $t^2_{l,r}$ is 15 minutes and between $t^1_{l',r}$ and $t^2_{l',r}$ is 12 minutes. This would lead to the situation in Figure 5.2, where a is the buffer time between $t^1_{l,r}$ and $t^1_{l',r}$ at r. In order to fit $t^1_{l',r}$ and $t^2_{l',r}$ between $t^1_{l,r}$ and $t^2_{l,r}$, a has to be strictly smaller than three. So, the smallest buffer time between a train of line l and line l' at r is smaller than or equal to one-and-a-half minutes, which is much smaller than the six minutes in case $f_l = f_{l'} = 5$.

The shared infrastructure resource that is mostly referred to in this chapter, is a station area.

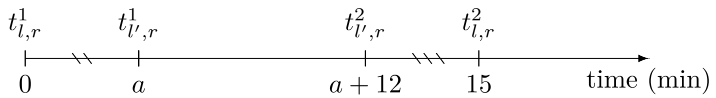


Figure 5.2: If lines l and l' have frequencies $f_l = 4$ and $f_{l'} = 5$ respectively, then once in 60 minutes two trains ($t^1_{l',r}$ and $t^2_{l',r}$) of line l' will pass in between two trains ($t^1_{l,r}$ and $t^2_{l,r}$) of line l at shared infrastructure resource r (pigeon hole principle). Without loss of generality we can assume that this happens in the first quarter. Here $a \in \mathbb{R}$ and $0 < a < 3$.

The best value of the minimum buffer time between two lines at a shared infrastructure resource is bounded by the following formula: *If trains that operate on a line are equally spread over the period of the cyclic timetable, the best value of the minimum buffer time between line l and line l' with frequencies $f_l \leq f_{l'}$ respectively, on a shared infrastructure resource, is smaller than or equal to ($\leq$)*

$$\frac{\frac{P}{f_l} - (\lceil \frac{f_{l'}}{f_l} \rceil - 1)\frac{P}{f_{l'}}}{2} \tag{5.3}$$

where P is the period length of the cyclic timetable and $\lceil x \rceil$ equals the smallest integer y with $y \geq x$.

Proof. Let r be a shared infrastructure resource of line l and l'. Without loss of generality, we assume occupation intervals of length zero, since any larger occupation interval will simply induce smaller buffer times. By the pigeon hole principle, there are two trains of line l in between which $\lceil f_{l'}/f_l \rceil$ trains of line

l' are passing at r. With the same notation as in the example above, we have, in chronological order, $t^1_{l,r}, t^1_{l',r}, \cdots, t^{\lceil f_{l'}/f_l \rceil}_{l',r}$ and $t^2_{l,r}$. The spread between trains $t^1_{l,r}$ and $t^2_{l,r}$ equals P/f_l minutes and between trains $t^1_{l',r}$ and $t^{\lceil f_{l'}/f_l \rceil}_{l',r}$ equals $(\lceil f_{l'}/f_l \rceil - 1)P/f_{l'}$ minutes. So, the buffer time between $t^1_{l,r}$ and $t^1_{l',r}$ plus the buffer time between $t^{\lceil f_{l'}/f_l \rceil}_{l,r}$ and $t^2_{l',r}$ equals $P/f_l - (\lceil f_{l'}/f_l \rceil - 1)P/f_{l'}$. Thus at least one of these two buffer times is smaller than half of this value, which is the bound given in (5.3). □

If the upper bound in (5.3) is strictly smaller than the minimum necessary buffer time according to safety regulations in the network, then line l with frequency f_l and line l' with frequency $f_{l'}$ are not feasible together. In the example, if the minimum necessary buffer time according to safety regulations is two minutes, then these lines l and l' cannot be combined at frequencies $f_l = 4$ and $f_{l'} = 5$.

Proof of Claim 1. We first show that expression (5.3) is bounded above by $P/(2f_{l'})$. Thereafter, we show that this upper bound is a tight bound in case f_l equals $f_{l'}$ or $f_{l'}$ is a multiple of f_l. As a result, it follows that the upper bound on the minimum buffer time between two lines l and l' with frequency f_l and $f_{l'}$ on a shared infrastructure resource is no greater than for two lines at the higher frequency of f_l and $f_{l'}$. We can write

$$f_{l'} = \alpha f_l + \beta, \tag{5.4}$$

with $\alpha, \beta \in \mathbb{Z}^+$ and $\beta < f_l$. Then we have:

$$\begin{aligned}
\frac{\frac{P}{f_l} - (\lceil \frac{f_{l'}}{f_l} \rceil - 1)\frac{P}{f_{l'}}}{2} &= \frac{P}{2f_{l'}}\left(\frac{f_{l'} - (\lceil \frac{f_{l'}}{f_l} \rceil - 1)f_l}{f_l}\right), \\
&= \frac{P}{2f_{l'}}\left(\frac{\alpha f_l + \beta - (\lceil \frac{\alpha f_l + \beta}{f_l} \rceil - 1)f_l}{f_l}\right) \\
&= \frac{P}{2f_{l'}}\left(\frac{\alpha f_l + \beta - (\alpha + \lceil \frac{\beta}{f_l} \rceil - 1)f_l}{f_l}\right) \\
&= \frac{P}{2f_{l'}}\left(\frac{\beta - \lceil \frac{\beta}{f_l} \rceil f_l + f_l}{f_l}\right) \\
&\leq \frac{P}{2f_{l'}}.
\end{aligned} \tag{5.5}$$

Formula (5.3) is equal to the upper bound $P/(2f_{l'})$ in case f_l equals $f_{l'}$ or $f_{l'}$ is a multiple of f_l ($f_{l'} = kf_l$, $k \in \mathbb{Z}_0^+$):

$$\frac{\frac{P}{f_l} - (\lceil \frac{kf_l}{f_l} \rceil - 1)\frac{P}{kf_l}}{2} = \frac{\frac{P}{f_l} - (k-1)\frac{P}{kf_l}}{2} = \frac{kP - (k-1)P}{2kf_l} = \frac{P}{2kf_l} = \frac{P}{2f_{l'}}. \tag{5.6}$$

□

On the one hand, Claim 1 and formula (5.3) can be used to find an upper bound on the minimum buffer time. On the other hand, they can be used to identify infeasible frequency combinations, or in other words to identify incompatible lines. We now propose a formula that lists incompatible frequencies for a certain line l with frequency f_l and a given minimum necessary buffer time B. Unfortunately, we have no formula that lists feasible frequency combinations given a minimum necessary buffer time. However, to be complete, we thereafter explain how a definitive answer can be obtained on the compatibility of two lines l and l' with frequencies f_l and $f_{l'}$.

Claim 2. *Suppose that the minimum necessary buffer time is $B > 0$ and P is the period length of the cyclic timetable. Then it holds that a line l with frequency $f_l < P/B$, is not compatible with lines l' for which the frequency*

$$f_{l'} \in \left] \frac{P}{\frac{kP}{f} + 2B}, \frac{P}{\frac{kP}{f} - 2B} \right[\cap \mathbb{Z} \setminus \{\frac{f_l}{k}\}, \tag{5.7}$$

for a (certain) $k \in \{1, \cdots, f_l\}$, on shared infrastructure resources.

Proof. Without loss of generality, we still assume occupation intervals of length zero, since any larger occupation interval will induce smaller buffer times.

First, suppose $f_{l'} \in \left]\frac{P}{\frac{kP}{f_l}+2B}, \frac{P}{\frac{kP}{f_l}}\right[\cap \mathbb{Z}$ with $k \in \{1, \cdots, f_l\}$. Then $f_{l'} < \frac{P}{\frac{kP}{f_l}}$, thus $kf_{l'} < f_l$ or equally $P/f_{l'} > kP/f_l$. This implies that somewhere in the cyclic timetable, k trains of line l will pass between two trains of line l' at shared infrastructure resources r. Consider function $fun(x) = \frac{P}{\frac{kP}{f_l}+x}$, which is continuous and decreasing for $x \in [0, 2B]$. According to the intermediate value theorem, there exists a $B' \in]0, 2B[$ such that $fun(B') = f_{l'}$. Thus $f_{l'} = \frac{P}{\frac{kP}{f_l}+B'} = \frac{P}{kP+B'f_l}f_l$. Using the same notation as before, the buffer time between $t^1_{l',r}$ and $t^1_{l,r}$ plus the buffer time between $t^k_{l,r}$ and $t^2_{l',r}$ equals

$$\frac{P}{f_{l'}} - \frac{kP}{f_l} = \frac{kP + B'f_l}{f_l} - \frac{kP}{f_l} = B' > 0. \tag{5.8}$$

Since $0 < B' < 2B$, there is at least one of the two buffer times that is not larger than the minimum necessary buffer time B. Thus line l and l' are not compatible, i.e. their frequencies f_l and $f_{l'}$ cannot be combined if they share infrastructure.

Secondly, suppose $f_{l'} \in]\frac{P}{\frac{kP}{f_l}}, \frac{P}{\frac{kP}{f_l}-2B}[\cap \mathbb{Z}$ with $k \in \{1, \cdots, f_l\}$. Then $f_{l'} > \frac{P}{\frac{kP}{f_l}}$, thus $kf_{l'} > f_l$ or equally $P/f_{l'} < kP/f_l$. This implies that somewhere in the cyclic timetable, two trains of line l' will pass between a certain train of line l and k trains later of this line l at a shared infrastructure resource r. For a visual representation of the latter implication, see Figure 5.3. Consider function $fun(x) = \frac{P}{\frac{kP}{f_l}-x}$, which is continuous and increasing for $x \in [0, 2B]$. According to the intermediate value theorem, there exists a $B' \in]0, 2B[$ such that $fun(B') = f_{l'}$. Thus $f_{l'} = \frac{P}{\frac{kP}{f_l}-B'} = \frac{P}{kP-B'f_l}f_l$. The buffer time between $t^1_{l,r}$ and $t^1_{l',r}$ plus the buffer time between $t^2_{l',r}$ and $t^k_{l,r}$ equals

$$\frac{kP}{f_l} - \frac{P}{f_{l'}} = \frac{kP}{f_l} - \frac{kP - B'f_l}{f_l} = B' > 0. \tag{5.9}$$

Since $0 < B' < 2B$, there is at least one of the two buffer times that is not larger than the minimum necessary buffer time. Thus line l and l' are not compatible, i.e. their frequencies f_l and $f_{l'}$ cannot be combined if they share infrastructure. □

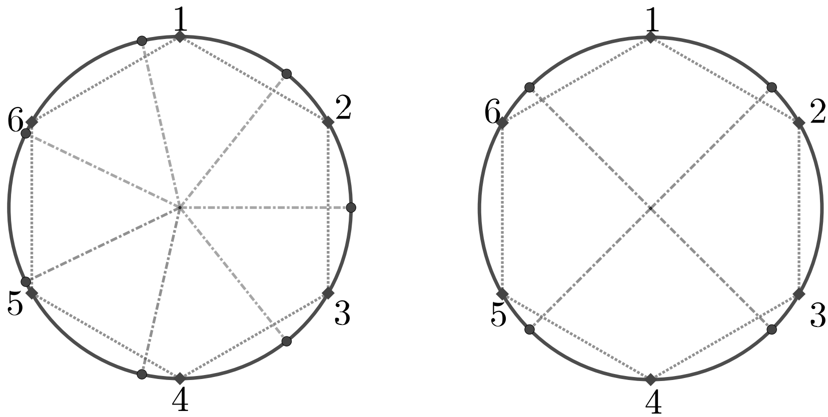
Example Suppose B equals 2, P equals 60, f_l equals 5 and k equals 1, then

$$\left] \frac{P}{\frac{kP}{f_l} + 2B}, \frac{P}{\frac{kP}{f_l} - 2B} \right[= \left] \frac{60}{12+4}, \frac{60}{12-4} \right[=]3.75, 7.5[. \tag{5.10}$$

This implies that $f_l = 5$ is not compatible with frequencies $f_{l'} \in \{4, 6, 7\}$.

It is also easy to see that if frequency f_l is not compatible with $f_{l'}$ that f_l will also not be compatible with multiples of $f_{l'}$.

As mentioned earlier, the above formulas only give a definitive answer if two frequencies are not compatible. However, in case one of both frequencies f_l and $f_{l'}$ is a multiple of the other frequency, e.g. $f_{l'} = kf_l$, Claim 1 determines a tight (upper) bound on their minimum buffer time (assuming occupation intervals of length zero). If $P/2f_{l'}$ exceeds B, with B the minimum necessary buffer time, both frequencies are compatible. Furthermore, if the above tests cannot give a definitive answer on the question whether two frequencies are compatible on a shared infrastructure resource, the following test can give a definitive answer.



(a) However you place the trains of lines l and l', there will always be two successive trains of line l, i.e. 'a train and ($k=$)1 train later of line l', in between which two trains of line l' pass the shared infrastructure. In the current configuration the two involved trains of line l are train 5 and 6.

(b) However you place the trains of lines l and l'', there will always be 'a train and ($k=$)2 trains later' of line l in between which two trains of line l'' pass the shared infrastructure. In the current configuration this statement holds for train 3 and 5, but also for train 6 and 2.

Figure 5.3: Example: If $f_l = 6$, $P = 60$ and $B = 3$, then $f_{l'} = 7 \in \left]\frac{P}{\frac{kP}{f_l}}, \frac{P}{\frac{kP}{f_l}-2B}\right[\cap \mathbb{Z}$ for $k=1$ and $f_{l''} = 4 \in \left]\frac{P}{\frac{kP}{f_l}}, \frac{P}{\frac{kP}{f_l}-2B}\right[\cap \mathbb{Z}$ for $k=2$. The big circle represents a cyclic time axis. The indications on the big circle indicate when the trains of line l (diamonds), l' (circles, left figure) and l'' (circles, right figure) pass some shared infrastructure. The trains of line l are numbered. The fact that the trains of a line are equally spread over the period of the timetable is shown with the dotted lines.

Claim 3. *Lines l and l' with frequencies f_l and $f_{l'}$ respectively, that share infrastructure, are compatible in a timetable with period P and minimum necessary buffer time B with $0 \leq 2B < P/f_l$ if and only if*

$$k'P/f_{l'} \in \bigcup_{k \in \{0,\cdots,f_l-1\}} [kP/f_l, (k+1)P/f_l - 2B] \quad \forall k' \in \{0, \cdots, f_{l'}-1\}. \tag{5.11}$$

Proof. First it is shown that if there exists a compatible disposition for the trains of lines l and line l' in a timetable with period P, then there also exists a compatible disposition where the minimum buffer time between the two lines exactly equals B. Let M be the minimum buffer time between lines l and l'. Since l and l' are assumed to be compatible, it holds that $M \geq B$. Without loss of generality, we can assume that the trains of line l pass at time instants kP/f_l for $k \in \{0, \cdots, f_l - 1\}$ and the trains of line l' pass at time instants $k'P/f_{l'} + B'$ for $k' \in \{0, \cdots, f_{l'} - 1\}$ and with $B \leq B' \leq P/f_l - B$. As in the other proofs in this section, we also assume occupation intervals of length zero. Then, M equals the following formula:

$$M = \min\{|kP/f_l - (k'P/f_{l'} + B')| \mid \forall k \in \{0, \cdots, f_l - 1\}, \forall k' \in \{0, \cdots f_{l'} - 1\}\}. \tag{5.12}$$

Note that $M \leq B'$ and we already mentioned that $M \geq B$. Now subtract the value $M - B$ from the time instants $k'P/f_{l'} + B'$ for line l'. Then, line l and l' are still compatible, since the buffer times between trains from line l and l' are still larger than or equal to B and by construction, at least one of them will be equal to B.

Now, note that for each $k' \in \{0, \cdots, f_{l'} - 1\}$, there exists a $k \in \{0, \cdots, f_l - 1\}$ such that $kP/f_l \leq k'P/f_{l'} + B' - (M - B) \leq (k+1)P/f_l$. Furthermore, it holds that

$$(k+1)P/f_l - (k'P/f_{l'} + B' - (M - B)) \geq B, \tag{5.13}$$

such that

$$(k+1)P/f_l - k'P/f_{l'} \quad \geq \quad 2B + B' - M \tag{5.14}$$

$$\geq \quad 2B. \tag{5.15}$$

Since k' was randomly chosen, formula (5.11) holds. This proves the first implication.

Now suppose no longer that l an l' are compatible, but that $k'P/f_{l'} \in \bigcup_{k \in \{0, \cdots, f_l - 1\}} [kP/f_l, (k+1)P/f_l - 2B]$ for all $k' \in \{0, \cdots, f_{l'} - 1\}$. Plan the trains of line l at time instants kP/f_l with $k \in \{0, \cdots, f_l - 1\}$. Plan the trains of line l' at time instants $(k'P/f_{l'} + B)$ with $k' \in \{0, \cdots, f_{l'} - 1\}$. Let k' be an element of $\{0, \cdots, f_{l'} - 1\}$. The assumption then implies that there exists a $\tilde{k}$ such that $k'P/f_{l'} \in [\tilde{k}P/f_l, (\tilde{k}+1)P/f_l - 2B]$. Thus, there exists an $0 \leq A \leq P/f_l - 2B$ such that $k'P/f_{l'} = \tilde{k}P/f_l + A$. Furthermore, this implies that the train of line l' uses the shared infrastructure between the trains at time instants $\tilde{k}P/f_l$ and $(\tilde{k}+1)P/f_l$ of line l. The buffer times between these train pairs are

$$\begin{aligned}
\frac{k'P}{f_{l'}} + B - \frac{\tilde{k}P}{f_l} &= \Big(\frac{\tilde{k}P}{f_l} + A\Big) + B - \frac{\tilde{k}P}{f_l} \\
&= A + B \\
&\geq B, \qquad (5.16)
\end{aligned}$$

$$\begin{aligned}
\frac{(\tilde{k}+1)P}{f_l} - (\frac{k'P}{f_{l'}} + B) &= \frac{(\tilde{k}+1)P}{f_l} - \Big(\Big(\frac{\tilde{k}P}{f_l} + A\Big) + B\Big) \\
&= \frac{P}{f_l} - A - B \\
&\geq \frac{P}{f_l} - (\frac{P}{f_l} - 2B) - B = B. \qquad (5.17)
\end{aligned}$$

Thus the buffer time between both train pairs exceeds the minimum necessary buffer time. Since this holds for any k', this implies that the two lines are indeed compatible. This proves the other implication of the claim. □

Claim 3 can be used to make a list of frequencies that are compatible with a given frequency.

5.3 Methodology

This section proposes an integrated approach that constructs a line plan from scratch that minimizes a weighted sum of operator and passenger cost and allows a feasible and robust timetable. First a timetable-feasible line plan is constructed. Then, iteratively and interactively, a line planning module produces a line plan, and for that line plan, a timetabling module produces a timetable that maximizes the (minimum) buffer times between train pairs. In each iteration an analysis of the timetable indicates how the line plan could be adapted in order to allow for a more robust timetable. This adaptation increases the flexibility of the line plan which is used, in the timetabling module, to increase the minimum buffer times. The line planning module then calculates a new line plan that includes this adaptation while minimizing the weighted sum of operator and passenger costs. This feedback loop stops when there is no further improvement possible or if there is no improvement for the minimum buffer times between train pairs observed during a certain number of iterations.

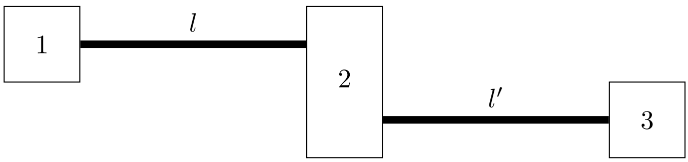
We first discuss the line planning module and the timetabling module separately and then the integration of both. Both the timetabling and the line planning module consist of an exact optimization model, though our combined approach, and the fact that we do not always solve the models to optimality, result in an overall heuristic method.

5.3.1 Line planning module

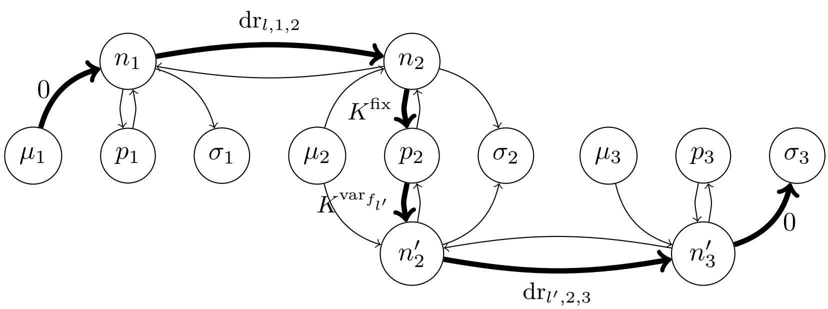
Constructing a line plan consists of selecting, from a pool of predetermined lines, a set of lines which meet certain requirements. The line pool is not exhaustive; there are many more lines possible than those considered, but the set is reduced to those that meet certain criteria, as discussed with the rail operator. This also keeps the problem size small. The model performs three functions: (i) selecting the lines and frequencies and creating a valid plan, (ii) routing passengers between their station of origin and their destination station and (iii) relating passenger routes to line selections.

Let us first define the set of all lines available: $\mathcal{L}$. For every line $l \in \mathcal{L}$, we define a set of valid frequencies for the line: $\mathcal{F}_l$. The operator must meet certain obligations for any valid line plan and may not exceed certain operational limits. These restrictions are referred to as service constraints. We define all of these in terms of a set of resources $\mathcal{R}$, and define all limitations as either a minimum (rmin_r) or maximum (rmax_r) number of trains using that resource $r \in \mathcal{R}$ every period. The subset of lines that makes use of resource $r \in \mathcal{R}$ is defined as $\mathcal{L}_r$. Let $c_{l,f}$ be the cost to the operator for operating line l at frequency f.

The line planning module starts from a known origin-destination (OD) matrix containing the passenger demand for traveling between every origin and destination, where origins and destinations are simply stations in the rail network. Let $\mathcal{S}$ be the set of stations. For every two stations $s_1, s_2 \in \mathcal{S}$ the demand d_{s_1,s_2} is known. In a graph, the passengers are modeled as a flow from each station of origin to every relevant destination station. The structure of this graph (nodes and edges) is uniquely determined (i) by the network (stations and station links) and (ii) by the lines considered in the line pool. Furthermore, this graph captures the passenger cost in terms of drive time on lines and estimated transfer time between lines in case a transfer is required (estimated based on the frequency). We refer to this graph as the *passenger graph.* We now explain the construction of this passenger graph in more detail. An example can be found in Figure 5.4 for a network with three stations, 1, 2 and 3, and two lines l and l' visiting two of the stations each. A passenger graph contains a *(line, frequency, station)* vertex for every line, frequency, and every station visited by that line. The edges of this graph represent travel possibilities, with the edge cost being the known train driving time or the estimated transfer time. Additionally, for every station s we have a platform vertex p_s with edges from and to every *(line, frequency, s)* vertex, where the costs correspond to an estimate of perceived transfer time, which consists of a fixed penalty component and a frequency-dependent component. Finally, this graph contains source μ_s and sink σ_s vertices for every station s where passengers originate from or terminate their travel. These vertices are connected to the appropriate *(line, frequency, station)* vertices with edges representing boarding or alighting from a line. These edges have zero cost. Source and sink vertices are modeled separately to ensure that line-to-line transfers are only possible via the platform



(a) Line structure (two lines, three stations).



(b) Corresponding passenger graph structure.

Figure 5.4: The upper figure shows a simple network with three stations 1, 2 and 3, and two lines l and l'. Line l visits stations 1 and 2, and line l' visits stations 2 and 3. Each line operates at just a single frequency ($f_l, f_{l'} \in \mathbb{Z}^+$). The lower figure shows the corresponding passenger graph structure used for this network. The notation has been simplified to keep the figure clear: node n_1 and n_2 represent node $(l, f_l, 1)$ and $(l, f_l, 2)$ respectively and node n'_2 and n'_3 represent $(l', f_{l'}, 2)$ and $(l', f_{l'}, 3)$ respectively. Costs are labeled on the edges for a passenger traveling from station 1 to station 3, transferring lines at station 2, with used edges in bold. The costs to the passenger are $\mathrm{dr}_{l,1,2}$, traveling (**dr**iving) on line l from station 1 to 2; fixed cost K^{fix} for a transfer and an additional $K^{\mathrm{var}} f_{l'}$ frequency dependent cost for transferring to line l'; and $\mathrm{dr}_{l',2,3}$ traveling on line l' from station 2 to 3.

vertex incurring the frequency-dependent costs.

Let V and E be the set of all vertices and edges of this graph, respectively, and τ_e be the cost to a single passenger of using edge $e \in E$. In total there are five types of edges.

- Type 1. From (l, f, s) to (l, f, s') for all lines $l \in \mathcal{L}$, $f \in \mathcal{F}_l$ and s and s' two successive stations visited by line l.
- Type 2. From (l, f, s) to p_s for all lines $l \in \mathcal{L}$, $f \in \mathcal{F}_l$ and s a station visited by line l and p_s the platform vertex of station s.
- Type 3. From p_s to (l, f, s) for all lines $l \in \mathcal{L}$, $f \in \mathcal{F}_l$ and s a station visited by line l and p_s the platform vertex of station s.
- Type 4. From μ_s to (l, f, s) for all lines $l \in \mathcal{L}$, $f \in \mathcal{F}_l$ and s a station visited by line l and μ_s the source vertex of station s.
- Type 5. From (l, f, s) to σ_s for all lines $l \in \mathcal{L}$, $f \in \mathcal{F}_l$ and s a station visited by line l and σ_s the sink vertex of station s.

This graph is similar to the *change&go* graph of Schöbel and Scholl (2006), but distinguishes between line transfers that, in our case, happen to lines with discrete frequencies, with a frequency-dependent cost. A more complex example with multiple frequencies per line can be found in Appendix A.2.

Let l_e be the line to which edge e of the passenger graph is related to and f_e be the frequency of the line to which e is related to. This line and frequency of an edge are uniquely defined, as the two vertices connected by edge e are either both related to the same line and frequency or only one of them is related to a line and frequency.

Let a_v^s be the flow of passengers originating from station s that enters vertex v minus the flow of passengers originating from station s leaving vertex v, where v is a vertex of the passenger graph. For vertices v of type *(line, frequency, station)* or platform vertices, a_v^s equals 0 for all stations $s \in \mathcal{S}$. All passengers that enter such a vertex, also leave again. For vertices v which are source vertices for a certain station s, passengers only leave to other stations according to the demand: $a_v^s = -\sum_{s' \in \mathcal{S}} d_{s,s'}$. For vertices v which are sink vertices for a certain station s, passengers coming from other stations s' are only entering: $a_v^{s'} = d_{s',s}$ for all stations $s' \in \mathcal{S}$.

For relating passengers to lines, let C_f be the passenger capacity of any line operating at frequency f. We are therefore assuming the same rolling stock unit type and sequence for every line, but a higher frequency provides more seats than a lower frequency. We require that not more passengers use a line as the line capacity permits for the frequency the line is operating at.

We use two classes of decision variables: $x_{l,f} \in \{0,1\}$ is a binary decision variable indicating whether or not line l is selected at frequency f, and y_s^e decides what number of passengers from station of origin s will use edge e in the passenger graph.

The line planning model is:

$$\text{Minimize} \quad \lambda \sum_{l \in \mathcal{L}} \sum_{f \in \mathcal{F}_l} c_{l,f} x_{l,f} + (1-\lambda) \sum_{e \in E} \sum_{s \in \mathcal{S}} \tau_e y_s^e \tag{5.18}$$

$$\text{s.t.} \quad \sum_{f \in \mathcal{F}_l} x_{l,f} \leq 1 \qquad \forall l \in \mathcal{L} \tag{5.19}$$

$$\sum_{l \in \mathcal{L}_r} \sum_{f \in \mathcal{F}_l} f x_{l,f} \geq \mathrm{rmin}_r \qquad \forall r \in \mathcal{R} \tag{5.20}$$

$$\sum_{l \in \mathcal{L}_r} \sum_{f \in \mathcal{F}_l} f x_{l,f} \leq \mathrm{rmax}_r \qquad \forall r \in \mathcal{R} \tag{5.21}$$

$$\sum_{(u,v) \in E} y_s^{(u,v)} - \sum_{(v,w) \in E} y_s^{(v,w)} = a_v^s \qquad \forall s \in \mathcal{S}, \forall v \in V \tag{5.22}$$

$$\sum_{s \in \mathcal{S}} y_s^e \leq C_{\mathrm{f}_e} x_{l_e,\mathrm{f}_e} \qquad \forall e \in E \tag{5.23}$$

$$x_{l,f} \in \{0,1\} \qquad \forall l \in \mathcal{L}, f \in \mathcal{F}_l \tag{5.24}$$

$$y_s^e \in \mathbb{R}^+ \qquad \forall s \in \mathcal{S}, e \in E. \tag{5.25}$$

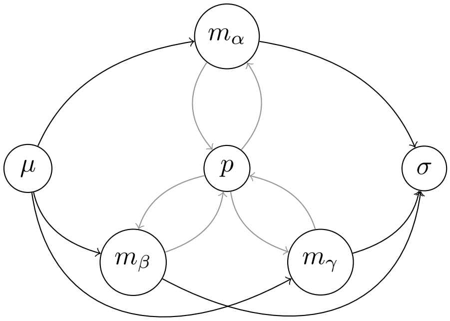
The objective function (5.18) is a weighted sum of the operator cost and the passenger travel time (drive time and transfer time), using a parameter $\lambda \in [0,1]$ to determine the relative importance of both components. The smaller λ the more important the passenger travel time compared to the operator cost and vice versa.

Constraints (5.19) ensure that a line is chosen with at most one frequency, i.e. combinations of frequencies for the same line are not permitted, as if this would be allowed, a discrete frequency (being the sum of the frequencies in the combination) would be present in the frequency set $\mathcal{F}_l$ for the line. Constraints (5.20) and (5.21) ensure that the obligatory and operational requirements are met for the line plan. Constraints (5.22) consist of the flow conservation constraints. There needs to be a flow of passengers from each station of origin with the appropriate arrive at every destination station, while the flow is conserved in every station. Constraints (5.23) link the flows of passengers to the line decisions. The presence of a positive passenger flow on an edge in the graph is dependent on some line being present in the plan. The maximum flow on that edge depends on the passenger capacity of the corresponding line at the appropriate frequency. Finally, constraints (5.24) and (5.25) restrict the line variables and flow variables to be binary variables and positive otherwise unrestricted variables, respectively.

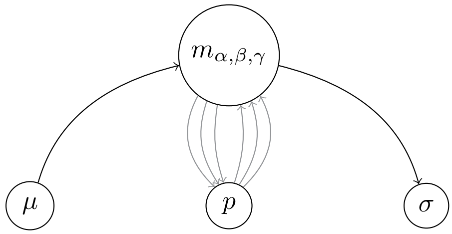
The model requires $|E||\mathcal{S}|$ flow decision variables, which is large due to the many edges in the described passenger graph. However, we observe that many of the vertices and edges in the graph are very similar and differ only in line frequency. For lines with many possible frequencies, there is significant duplication. For the edges related to a transfer at a station, the frequency is required to determine the cost to the passenger. However for all other edges the frequency information is redundant. Indeed, for one passenger, the cost of traveling on a line between stations does not depend on the frequency of that line. A first simplification of the model is that for each line and its frequencies, we replace the edges (and vertices) which do not depend on frequency with an edge (and vertex) only related to *line* and *station* instead of *line*, *frequency* and *station*. This is shown in Figure 5.5. The capacity of the replacement edge (and resulting right hand side of constraints (5.23)), is given by $\sum_{f \in \mathcal{F}_l} C_f x_{l_e,f}$.

Figure 5.5 shows the graph structure for a single station and a single line with three frequencies as originally described (Figure 5.5a) and with the explained reductions (Figure 5.5b). Nodes μ and σ are respectively the station source and sink vertices for passengers and p is the platform vertex for that station. The vertices m_α, m_β, m_γ are the *(line, frequency, station)* vertices for the three considered frequencies of the line, in that station. The gray edges are the transfer edges (though no other lines are shown). Edges connecting these vertices m_α, m_β, and m_γ to corresponding vertices at other stations are not shown. Vertex $m_{\alpha,\beta,\gamma}$ is the combination of the vertices m_α, m_β, and m_γ. The edge between μ and $m_{\alpha,\beta,\gamma}$, and between $m_{\alpha,\beta,\gamma}$ and σ, is the combination of the edges between μ and m_α, m_β, and m_γ in (Figure 5.5a), and m_α, m_β, and m_γ and σ, respectively. In Appendix A.2 a more complex example of a passenger graph reduction can be found.

A second simplification of the model is that we consider transfer edges only at a minimal set of transfer stations. This set of stations is fixed beforehand and suffices to facilitate all optimal passenger flows, when every passenger's origin-destination pair is considered individually. Any solution that is feasible for this restricted problem is feasible if transfer edges are included for any



(a) Original graph structure.



(b) Reduced graph structure.

Figure 5.5: The full and reduced graph structure for a single line l with three frequencies, α, β and γ in $\mathbb{Z}^+$, at a single station s. The notation has been simplified to keep the figure clear: node m_i represents node (l, i, s) for frequency $i \in \{\alpha, \beta, \gamma\}$. Node $m_{\alpha,\beta,\gamma}$ is the replacement node of nodes m_α, m_β and m_γ in the reduced graph structure.

station, but some solutions that are feasible if transfers are permitted anywhere may not be feasible with the restriction (although we have not observed this). At stations where we do not permit transfers, we do not include transfer edges and this reduces the total number of edges in the graph by between 23% and 34% when tested for a range of line pools. Finally, we can determine that only a subset of all edges needs to be used for the flows from a given station of origin; generally it is never true that in an optimal solution passengers will be assigned an edge that travels ‘towards’ the station they originate from. This is a third way to simplify the model.

By making these three alterations, we find that the line planning problem is solvable directly as a MILP, though not to optimality in the time frame we require. When using the model (5.18)-(5.25) to construct a line plan without extra restrictions, a time limit of one hour is set, unless a gap smaller than 0.5% between the solution and the best lower bound is found earlier (in most cases the gap limit is reached, but for some weightings of the objectives, one hour is insufficient). For a reduced line pool that we use in the integrated approach described later, the problem becomes easier and is solvable to optimality in an acceptable time frame.

The formulation (5.18)–(5.25) defines the basic line planning model. However, when searching for line plans that only differ a little from a given line plan, we may impose some additional restrictions. The simplest types are the following:

$$\sum_{l\in\mathcal{L}}\sum_{f\in\mathcal{F}_l} f x_{l,f} \geq k_1 \tag{5.26}$$

$$\sum_{l\in\mathcal{L}}\sum_{f\in\mathcal{F}_l} f x_{l,f} \leq k_2 \tag{5.27}$$

That is, we require that the total number of (one-directional) trains running in the network per hour is between some upper and lower bound. This may be necessary, for example, to find solutions that do not differ too much from some original solution. The reason we do this is that, seen from the point of view of the timetabling module, two solutions that differ only in line frequency but not in line routes can be very different (think of the frequency combination infeasibility for example). Without the former constraints, when seeking a line plan that is similar but different to a given plan, a change of frequency would not maintain the similarities in timetabling we seek.

Now, suppose we are given a line plan or a partial line plan, in the form $\mathcal{X} = \{(l,f),(l',f'),(l'',f''),...\}$ where every (l,f) in $\mathcal{X}$ is a valid line and frequency combination and that this (partial) line plan should not be in the solution. Then we may impose the following constraint for each such line plan:

$$\sum_{(l,f)\in\mathcal{X}} x_{l,f} \leq |\mathcal{X}|-1. \tag{5.28}$$

Such constraints are used to forbid solutions we have already discovered and do not wish to find again, and also to forbid partial solutions that are already known to be problematic for timetabling, i.e. that lead to timetable-infeasibility. Finally, and in a similar way, we may have some given line plan $\mathcal{X}$ and desire that the solution line plan contains at least k lines from the plan:

$$\sum_{(l,f)\in\mathcal{X}} x_{l,f} \geq k. \tag{5.29}$$

Such constraints will ensure that a discovered line plan is similar to some previous line plan, while differing by some number of (unspecified) lines. If instead the lines that may differ are specified, we can fix the variables of the lines that need to remain unchanged and only permit those variables corresponding to the specified lines that are allowed to change (along with variables corresponding to lines not in the plan). These extra restrictions are used in the integrated approach when looking for a similar line plan that is more flexible, i.e. allows a more robust timetable.

5.3.2 Timetabling module

The timetabling module is based on a PESP model. We indicate our event-activity network as $(\mathcal{E}, \mathcal{A})$. The set of trains is indicated as T, the set of lines in the line plan (output of the line planning module) as $\mathcal{X}$, the line operated by train t is indicated as ℓ_t, the set of station areas is $\mathcal{S}$ and the set of station areas on a line l (independent of an actual stop in these stations) is indicated as $\mathcal{S}_l$. As we assume a railway network with limited shunt capacity, our model assumes that all the trains can and must turn on their platform at terminal stations. The set T_{turn} contains the train couples (t, t') for which it holds that t becomes train t' after turning on the platform in its terminal station. Trains t and t' share the same rolling stock. Line ℓ_t and $\ell_{t'}$ contain the same stations but in opposite direction. The set $T_{\text{line spread}}$ contains the train couples (t, t') where t and t' are two successive trains of the same line, i.e. no other train operating on the same line drives in between them.

The event set $\mathcal{E}$ of the event-activity network consists of the following events.

- The reservation of a station area s by a train t is a reservation event (t, s, res). We define $\mathcal{E}^{\text{res}}$ as $\{(t, s, \text{res}) \mid \forall t \in T, s \in S_{\ell_t}\}$.
- The release of a station area s by a train t is a release event (t, s, rel). We define $\mathcal{E}^{\text{rel}}$ as $\{(t, s, \text{rel}) \mid \forall t \in T, s \in S_{\ell_t}\}$.
- The reservation of a platform $\rho_{\tilde{s}_t,t}$ by a train t in its terminal station $\tilde{s}_t$ in order to turn, is a platform reservation event $(t, \rho_{\tilde{s}_t,t}, \text{res})$. We define $\mathcal{E}^{\text{res},p}$ as $\{(t, \rho_{\tilde{s}_t,t}, \text{res}) \mid \forall t \in T\}$.
- The release of a platform $\rho_{\tilde{s}_t,t}$ by a train t in its terminal station $\tilde{s}_t$ in order to turn, is a platform release event $(t, \rho_{\tilde{s}_t,t}, \text{rel})$. We define $\mathcal{E}^{\text{rel},p}$ as $\{(t, \rho_{\tilde{s}_t,t}, \text{rel}) \mid \forall t \in T\}$.

The following inclusions hold $\mathcal{E}^{\text{res},p} \subset \mathcal{E}^{res} \subset \mathcal{E}$ and $\mathcal{E}^{\text{rel},p} \subset \mathcal{E}^{\text{rel}} \subset \mathcal{E}$ and $\mathcal{E} = \mathcal{E}^{\text{res}} \cup \mathcal{E}^{\text{rel}}$. So platform $\rho_{\tilde{s}_t,t}$ of train t in its terminal station can be interpreted as an extra station where the train arrives after entering in its terminal station $\tilde{s}_t$. Note that the event set consists of station reservation and release times instead of the more common arrival and departure times in stations. From a macroscopic viewpoint these reservation and release times of a station area can be used to derive arrival and departure times on the platforms. Since we did not construct the timetable on the signaling level, we cannot fully guarantee that a timetable, that is feasible according to our model, is also conflict-free in practice on the microscopic level. However, all the timetables constructed during our case study that were checked by the railway operator, were found suitable to be implemented in practice.

The activity set $\mathcal{A}$ contains:

- *running activities* between the release of a train in a station and the reservation of this train of the next station on its line. Let $\mathcal{A}^{\text{run}} = \{((t, s, \text{rel}), (t, s', \text{res})) \in \mathcal{E}^{\text{rel}} \times \mathcal{E}^{\text{res}} \,|\, \forall t \in T$ and s and s' successive stations of $\ell_t\}$;
- *station activities* between the reservation and the release of a train in a station on its line. Let $\mathcal{A}^{\text{station}} = \{((t, s, \text{res}), (t, s, \text{rel})) \in \mathcal{E}^{\text{res}} \times \mathcal{E}^{\text{rel}} \,|\, \forall t \in T, s \in S_{\ell_t} \setminus \{\rho_{\tilde{s}_t,t}\}\}$;
- *turn activities* between the reservation and the release of a train on its platform in its terminal station. Let $\mathcal{A}^{\text{turn}} = \{((t, \rho_{\tilde{s}_t,t}, \text{res}), (t, \rho_{\tilde{s}_t,t}, \text{rel})) \in \mathcal{E}^{\text{res},p} \times \mathcal{E}^{\text{rel},p} \,|\, \forall t \in T\}$;
- *buffer activities* between the release of one train and the reservation of another train in the same station area. Let $\mathcal{A}^{\text{buffer}} = \{((t, s, \text{rel}), (t', s, \text{res})) \in \mathcal{E}^{\text{rel}} \times \mathcal{E}^{\text{res}} \,|\, \forall t, t' \in T : t \neq t', s \in S_{\ell_t} \cap S_{\ell'_t}\}$;
- *line spread activities* between the reservations of two successive trains on the same line in the stations on their line. Let $\mathcal{A}^{\text{line spread}} = \{((t, s, \text{res}), (t', s, \text{res})) \in \mathcal{E}^{\text{res}} \times \mathcal{E}^{\text{res}} \,|\, \forall t, t' \in T : (t, t') \in T_{\text{line spread}}, s \in S_{\ell_t}\}$;
- *turn connection activities* between the release of a train of the platform in its terminal station and the release of the next train of the opposite line that leaves from that station area. Let $\mathcal{A}^{\text{turn-con}} = \{((t, \rho_{\tilde{s}_t,t}, \text{rel}), (t', \tilde{s}_t, \text{rel})) \in \mathcal{E}^{\text{rel},p} \times \mathcal{E}^{\text{rel}} \,|\, \forall t, t' \in T : (t, t') \in T_{\text{turn}}\}$. This next train is the same physical train.

The objective of the timetabling model is to maximize the minimum buffer times between train pairs. In terms of the event-activity graph, this is equal to maximizing the minimum activity time of the buffer activities. Mathematically we have

$$\max \min_{a=(i,j)\in\mathcal{A}^{\text{buffer}}} (\pi_j - \pi_i + k_a P), \tag{5.30}$$

where π_i and π_j are the event times of event i and j respectively, which together define a buffer activity, and the terms $k_a P$ avoid negative buffer times. More specifically, k_a is a binary variable and P is the period length of the cyclic timetable. The objective function is not linear, but as it is a max-min objective function, it can easily be linearized by introducing the decision variable $z \in [0, P[$. This decision variable z represents the overall minimum buffer time. We add the constraints

$$z \leq \pi_j - \pi_i + k_a P \qquad \forall a = (i, j) \in \mathcal{A}^{\text{buffer}} \tag{5.31}$$

and we change the objective function to the maximization of z: $\max z$. The complete model is then the following.

$$\max \quad z \tag{5.32}$$

$$z \leq \pi_j - \pi_i + k_a P \qquad \forall a = (i,j) \in \mathcal{A}^{\text{buffer}}$$

$$L_a \leq \pi_j - \pi_i + k_a P \leq U_a \qquad \forall a = (i,j) \in \mathcal{A} \tag{5.33}$$

$$0 \leq \pi_i < P \qquad \forall i \in \mathcal{E} \tag{5.34}$$

$$0 \leq z < P \tag{5.35}$$

$$k_a \in \{0,1\} \qquad \forall a = (i,j) \in \mathcal{A}. \tag{5.36}$$

Constraints (5.33) bound all activity times from below and above. The term $k_a P$ avoids negative activity times. To ensure a unique value for k_a, the value of U_a has to be smaller than the period length P. The specific values of U_a and L_a are listed in Table 5.1 for all activities $a \in \mathcal{A}$.

Activity	L_a	U_a
$((t,s,\text{rel}),(t,s',\text{res})) \in \mathcal{A}^{\text{run}}$	$\text{run}_{\ell_t,s,s'}$	$\text{run}_{\ell_t,s,s'}$
$((t,s,\text{res}),(t,s,\text{rel})) \in \mathcal{A}^{\text{station}}$	$\text{occ}_{\ell_t,s}$	$\text{occ}_{\ell_t,s}$
$((t,\rho_{\tilde{s}_t,t},\text{res}),(t,\rho_{\tilde{s}_t,t},\text{rel})) \in \mathcal{A}^{\text{turn}}$	ntt_s	$\frac{P}{\varphi_{\rho_{\tilde{s}_t,t}}}$
$((t,s,\text{rel}),(t',s,\text{res})) \in \mathcal{A}^{\text{buffer}}$	0	$P - \text{occ}_{\ell_{t'},s} - \epsilon$
$((t,s,\text{rel}),(t',s,\text{res})) \in \mathcal{A}^{\text{buffer}} : s = \rho_{\tilde{s}_t,t} = \rho_{\tilde{s}_t,t'}$	0	$P - \frac{P}{\varphi_{\rho_{\tilde{s}_t,t}}} - \epsilon$
$((t,s,\text{res}),(t',s,\text{res})) \in \mathcal{A}^{\text{line spread}}$	$\frac{P}{f_{\ell_t}}$	$\frac{P}{f_{\ell_t}}$
$((t,\rho_{\tilde{s}_t,t},\text{rel}),(t',\tilde{s}_t,\text{rel})) \in \mathcal{A}^{\text{turn-con}}$	0	0

Table 5.1: Lower and upper bounds for the PESP constraints (5.33).

The running activity times are bounded by the time that a train of line l needs between the release of a station s and the reservation of the next station s', indicated as $\text{run}_{l,s,s'}$. The running time between the terminal station of a train and the platform in its terminal station is zero minutes. The station activity times are bounded by the time that is necessary and provided for a line l to occupy a station s, indicated as $\text{occ}_{l,s}$. This is the time between the reservation and release time of that station. The turn activity times are bounded by the necessary turn time in the terminal station s, which is indicated as ntt_s and the time at which the next train arrives on that platform. Trains making use of the same turn platform all get the same maximum time to stay on that platform which is equal to the period length of the cyclic timetable divided by the number of trains that turn on platform p. The number of trains that turn on platform p is indicated as φ_p. The buffer activity times have to be positive and smaller than $P - \text{occ}_{\ell_t,s} - \epsilon$ to ensure that occupation intervals do not overlap, independently of the order of both trains that will be assigned. On platforms in terminal stations the upper bound is smaller, because trains occupy the platform for a longer time, i.e. the upper bound in our model is $P - \frac{P}{\varphi_{\rho_{\bar{s}_t},t}} - \epsilon$. Before initializing the timetabling module, a check is necessary to determine whether too many trains are scheduled on one platform, i.e. $\frac{P}{\varphi_p} \geq \text{ntt}_s$ must be satisfied. If so, the trains have enough time for turning, otherwise the timetable will be infeasible. The value of ϵ depends on the time discretization. We use 0.1 minutes. In this model, trains of a line are equally distributed over the period, and therefore the line spread activity times have to be equal to the period length divided by the line frequency. The frequency of a line l is indicated as f_l. The turn connection activity times have to be equal to zero, ensuring that the 'turning' platform is freed if the next train leaves in the opposite direction.

5.3.3 Integrated approach

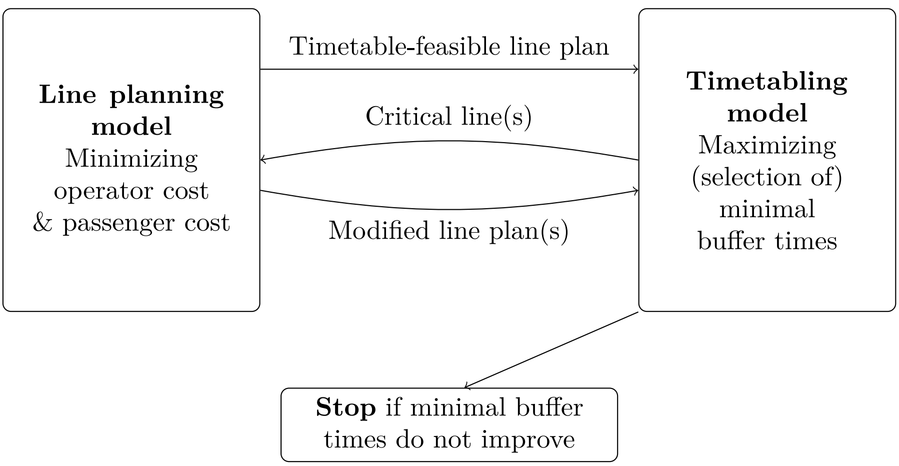


Figure 5.6: Overview of the integrated approach.

This section explains how the line planning and timetabling model can be integrated in one approach to construct a line plan and timetable that generate a low passenger and operator cost and maximize the buffer times between train pairs in order to provide a passenger robust railway schedule. The approach consists of a line planning and a timetabling module that work iteratively and interactively. The line planning module first creates an initial line plan, which is evaluated by the timetabling module. Based on the minimum buffer times between line pairs, a critical line in the line plan is identified. The line planning module then creates a new line plan with at least one different line, i.e. the time length of this critical line is changed. The goal is to create more flexibility in the line plan. This flexibility will be used by the timetabling module to improve its passenger robustness. This heuristic approach, which is divided into two parts, is now further explained. In Figure 5.6, a visual overview of the algorithm is presented and in Section 5.5.2 the approach is applied to an example.

Part 1: Initialization.

Step 1: Construct an initial line plan.
A line plan is constructed that satisfies the service constraints and optimizes a weighted sum of the passenger and operator cost with the line planning module. Beforehand, there is a check to eliminate infeasible lines from the line pool, as discussed in Section 5.2. Thereafter the timetabling module checks whether a feasible timetable can be constructed for this initial line plan. A feasible timetable is a timetable in which no occupation intervals of trains overlap: if a station or platform is occupied by one train, no other train can occupy this station or platform until the first train leaves it. In case the constructed line plan is not timetable-feasible, different strategies can be applied. A straightforward strategy is to take the second best line plan for the weighted sum of the passenger and operator cost and if the second best is not timetable-feasible then the third best and so on. The disadvantage of this strategy is that it is possible that a lot of line plans are to be tested before a timetable-feasible line plan is found, because no insight in the problem is used. We propose another more effective strategy for a network with restricted shunt capacity as is assumed in this research. Due to the restricted shunt capacity in the terminal stations, the occupation of the terminal stations is critical in finding a timetable-feasible line plan. So an effective strategy for looking for a timetable-feasible line plan with a close to optimal objective value restricts the number of lines that share a terminal station for example. If a line using a shared terminal station also passes a different station that may be a terminal station, a close to optimal solution is a line plan in which this line is replaced by one that ends at this alternative terminal station. This decreases the number of lines sharing a terminal station and in some cases has minimal impact on operator and passenger costs. This new line plan is only feasible in case all service constraints remain satisfied.

Part 2: Iterative steps.

Step 2: Evaluate the line plan.
Construct a timetable with the timetabling model that maximizes the minimum buffer times between a selection, or between all the train pairs in the line plan. Calculate the minimum buffer times between all line pairs in the line plan, and the overall minimum buffer time.
Test the following stopping criteria:

- STOP if the minimum buffer time is closer than 5% to the desired minimum buffer time. The *desired minimum buffer time* can be found by identifying the station area or track section which has the highest ratio of occupation time over free time (i.e. buffer times) and dividing the free time by the number of trains that pass by this section or station. Note that the desired minimum buffer time depends on the line plan. This stopping criterion is referred to as 'DES' (from *desired*).
- STOP if the minimum buffer times do not improve the best found value for three successive iterations. This stopping criterion is referred to as 'BFV' (from *best found value*).

Otherwise, select the most critical line from the list. The most critical line is the line that is responsible for the highest number of buffer times in the category of smallest buffer times in the list. This is illustrated in the example in Section 5.5.2. The thresholds to categorize the buffer times depend on the operator. In case of a tie, look at the next category of buffer times to identify the most critical line. If there is still a tie, let the decision be made by the line planning module in Step 3, based on the objective values there. Go to Step 3.

Step 3: Adapt the line plan by changing the stopping pattern.
Make a new line plan that alters the time length of the critical line by adding or removing a stop in a station on that line, such that this line becomes more flexible. This flexibility will be used to improve the buffer times in the timetabling module. This effect can be seen in the results and the example presented in Section 5.5. There are three important considerations. Firstly, changing the time length can also make a line infeasible, as discussed in Section 5.2, which has to be avoided. Secondly, an extra stop cannot be added to a line in cases where there are no skipped stations on the line. Thirdly, some stations cannot be skipped due to service constraints.

The line planning problem is potentially solved with three different line pools, sequentially, to attempt to find a feasible solution. If a feasible line plan is found, the line planning problem does not need to be solved for the other line pools in the sequence. The three line (and frequency) pools are as follows:

i. All lines of the solution of the previous iteration are fixed, including their frequency, except that of the critical line. We add all lines that differ by one stop from the critical line. For those lines we only allow the frequency of the critical line.
ii. All lines of the solution of the previous iteration are fixed, including their frequency, except that of the critical line. We add lines to the line pool that differ by one stop from the critical line, which we now allow at any frequency.
iii. Solution lines that share no stations with the critical line are fixed. We introduce lines that differ by one stop from the critical line and lines that differ from other non-fixed non-critical lines by one station, at any frequency.

Because the number of lines in the line pool and the number of feasible solutions is much more restricted, the run time for the line planning module is now much shorter. The objective function is the same as in Step 1. For the first line pool, if feasible, the best alternative line will be selected, i.e. the line that provides the lowest passenger and operator costs. For the second line pool, if feasible, one or more of these new lines will be selected, often with a frequency combination that sums to the frequency of the critical line. For the third line pool, one or more lines similar to the critical line will be selected, and other solution lines from the previous iteration may be replaced with one or more similar lines. A simple example of a solution from the third line pool is where a stop at a certain station is shifted from the critical line to a line that first skipped this station. The time length of the critical line changes by removing a stop and the station that is now skipped by the critical line is still served, but by another line. Note that in this example, the length of the non-critical line is also changed. A composition resulting from the second line pool is captured in the example in Section 5.5.2.

In the case a feasible solution is found, return to Step 2. In the case no feasible solution is found, and if there is a second most critical line, solve the three line planning problems for the second most critical line. Otherwise STOP.

End

The selected final solution is the combination of the line plan and the timetable constructed during the iterative approach, that results in the best minimal buffer time taken over all iterations. In case of a tie, the best weighted sum of passenger and operator cost is used as criterion. As a result, the selected final solution is always the best one found during the search. The intuition behind the integrated approach is the following. Changing the number of stops of a line changes the time length of the line. This time length of a line affects the flexibility of that line, e.g. Section 5.2.1. So we alter the stopping pattern of a line to make the line more flexible in order to improve the spreading in the whole network. The station where the stopping pattern is changed is decided upon by the line planning module, which takes a weighted sum of passenger and operator cost into account. These costs are not taken into account during the timetable construction. We indicate briefly that in general we do not require that the lines created to modify a line plan are all in the original pool of lines specified for the original problem. This explains why the adapted line plan can have a better weighted sum of passenger and operator cost than the original one. For the stopping criterion 'BFV' we take three non-improving iterations, to both restrict the run time while still allowing improvements that require multiple lines to change before a resultant improvement in minimal buffer time is observed.

5.4 Case study

The railway system on which the approach is tested is the S-tog network in Copenhagen, operated by the Danish railway operator DSB. This is a cyclic high-frequency network with a one hour period of repetition. This network transports 30 000 to 40 000 passengers per hour at peak times between 84 stations. The origin-destination data used comes from the operator, with non-zero demand for 65% of all possible pairings of stations. The network is visualized in Figure 5.7. It contains a central corridor, indicated in red; five 'fingers', indicated in blue; and a circle track, indicated in yellow. With almost no exception, there are at least two tracks in between every two adjacent stations and mostly exactly one track in each direction. We model the network with exactly one track in each direction between two stations and one platform in each direction in every station. One extra platform for turning is modeled in stations which can be used as intermediate terminal stations, see Figure 2.4c. The network is built such that crossings in between station areas are avoided by tunnels and bridges and there are only very few exceptions to this in the real network. Moreover, there are very few locations where trains from opposite directions have to cross each other during normal conditions. In this research we assume that trains in opposite directions only interact with each other in terminal stations. Furthermore, trains are not allowed to overtake.

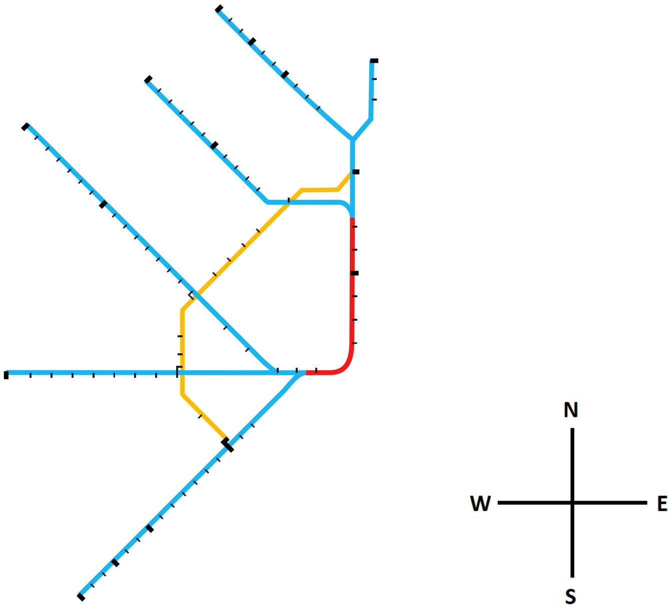


Figure 5.7: DSB S-tog network of Copenhagen.

Each train is modeled to occupy all station areas on its route for one minute. In case the train has a stop in the station area, this occupation time encompasses the dwell time, which is observed to be 20 seconds in practice. However, the occupation time is considered the same whether the train stops or not in this station area. This is to rule out that two trains or more can use this station area at the same time, while at most one of these has a stop. This is important, as we model only one platform in each station area and trains may not overtake. The buffer and occupation intervals of a station area are always disjoint and the buffer and occupation times count up to the length of the period. The driving times (an occupation time in a first station plus running time between this first station and a next station) are given by DSB S-tog and are independent of the line or train passing these stations. Only if a station is skipped, one minute is subtracted from the driving time between this station and the next station. Also this rule is inherited from DSB S-tog. In reality, partly the occupation time and partly the running time is superfluous, but we model this by subtracting one minute from the running time, in order to stick to the station occupations and in so-doing rule out overtaking conflicts.

During peak hour on weekdays, there is a service requirement of 30 trains per hour through the central corridor in each direction. The minimum *desired buffer time* (as defined in stopping criterion 'DES') in the DSB S-tog network is therefore one minute, which is (60 min - 30 min)/30, where 60 minutes is the period length of the cyclic timetable and 30 trains occupy a station in the central corridor each for one minute. The line pool originates from the operator and consists of 350 lines (170 if the frequencies are omitted), but actually, there exist 46 795 timetable-feasible lines that start and stop in a valid terminal station and never skip stations in the central corridor. One requirement specified by the operator is that only lines at frequency three, six, nine or twelve are allowed in the weekday line plan. This restriction decreases the probability of frequency combination infeasibility (though that is not necessarily the intention for the requirement). In order to enable and maintain the high frequency in the central corridor, the spreading of the trains in this part of the network is crucial. Therefore the timetabling module will be sequentially used twice with two different objective functions. First, the minimum buffer times in the central corridor are optimized. In a second optimization round, the minimum buffer times in the rest of the network are optimized while bounding the buffer times in the central corridor by the value found in the first optimization round. We also considered one combined weighted objective function, but this proved to be detrimental for the computation time of our experiments, i.e. the run times were significantly higher. The selected final solution is the combination of the line plan and the timetable constructed during the iterative approach, that has the best minimal buffer time in the central corridor as a first criterion, the best overall minimal buffer time in the network as a second criterion and the best weighted sum of passenger and operator cost as a third criterion, taken over all iterations. The second and third criterion are only used in case of a tie (in the first and second criterion, respectively).

We test our approach on ten line plans for this network. The approach can be applied to a pre-existing line plan, or applied to first create and then improve a line plan. The full approach is tested for five line plans created as described in Step 1 of the integrated approach, while the other five line plans come from the operator or are created by hand. The first two line plans (1-2) were recently in use for the S-tog network in Copenhagen. The current line plan is not considered as it is only temporarily active and specifically developed for implementing the new signaling system in the central corridor of the network. The third line plan (3) is a night line plan for weekdays. As the demand during night time is lower, the frequencies of the lines in this line plan are also lower. All other line plans are line plans that are planned with the requirements for use during daytime on weekdays. So, the setting of this third line plan is different from the other ones. This third line plan is also not the current plan in the S-tog network as at the present time a temporary plan is also in use during night time. The fourth up to the eighth line plan (4-8) are created within our algorithm by optimizing the weighted sum of the line planning model, using a range of weights that give distinct line plans. For each of these weights, the line planning model is solved with a one hour time limit and to a 0.5% relative gap limit, and terminate when either is reached. The line planning module is initially solved for finding distinct solutions with no consideration for the feasibility of timetables, except for infeasible lines as explained in Section 5.2. It is then tested whether or not the output line plans are timetable-feasible. For these considered weights, it appeared that only a single line plan (4) is feasible for timetabling. This endorses the statement that the output of a previous level in railway planning is not necessarily adequate for the next planning level (Schöbel, 2015). For those line plans that were not feasible, restrictions were introduced on the use of terminal platforms, requiring at most one line terminating at a station's terminal platform. This is described in Step 3 of the integrated approach in Section 5.3. This is sometimes too conservative, since it can be possible for more than one line to share a single terminal platform. Conversely this alone does not guarantee that a feasible timetable is present for a line plan, even though we observe that it is often a sufficient restriction. Applying this restriction, four other distinct line plans (5-8) are found. We note that, when considering the two line plan objectives of operator cost and passenger cost, none of the final four plans dominates any other. The ninth and the tenth line plan (9-10) are two special 'manually created' line plans, which are each based on one of the weighted-objective line plans (5 and 8 respectively). These paired plans only differ from the plan they are manually adapted from in their stopping pattern, as we force every line to stop in every station it passes, while the original line plans contain many skipped stations. We want to investigate whether each pair (5 and 9, 8 and 10) converges to a final line plan of similar quality when we modify stopping patterns of lines.

5.5 Results and discussion

In this section we show the results of the integrated approach for all ten line plans described in Section 5.4. Furthermore, we demonstrate the integrated approach for line plan 2 and include for this line plan the time-distance diagrams for the central corridor for the initial and the finally selected timetable. We end this section by showing the effect of the limited shunt capacity on the construction of a passenger robust timetable for the DSB S-tog network.

5.5.1 Results for ten line plans

A first performance indicator is the estimated operator cost of a line plan. This cost is calculated by the line planning module. The total operator cost of a line plan is simply the sum of the estimated operator costs for each line, which we take as given by the rail operator, here DSB. Each line in the pool has an operating cost associated with each frequency at which it could operate, and in calculating the total cost there are no additional considerations taken into account with respect to the combinations of lines.

A second performance indicator is the estimated passenger cost of a line plan. This cost is estimated by the line planning module. It is the sum of the travel times of all passengers in the origin-destination matrix. Because a timetable is not known (by the line planning module), the transfer time is estimated based on the frequency of the line as half the time in between two trains of that commuter line. For each passenger transfer an additional penalty of six minutes is added to the estimated passenger cost, as passengers perceive transfers worse than direct connections.

The third performance indicator is the minimum buffer time between train pairs in the central corridor of the DSB S-tog network, optimized by the timetabling module. The fourth performance indicator is the minimum buffer time between train pairs everywhere in the network, while bounding the minimum buffer time in the central corridor first. This fourth performance indicator is also optimized by the timetabling module. The focus on the minimum buffer time first in the central corridor of the network and thereafter on the overall minimum buffer time is in consultation with DSB S-tog.

A fifth performance indicator is the sum of the inverse of the minimum buffer times between train pairs in each station that they have in common (and pass by in the same direction). Maximizing the inverse minimum buffer times is interesting in order to give smaller buffer times a higher weight than larger buffer times. As in Dewilde et al. (2013) a buffer time smaller than the time discretization ϵ (here 0.1 minute) has a contribution of 15 to the sum of the inverse buffer times. So the lower the sum of the inverse buffer times the better, because this generally means larger buffer times. The results are summarized in Table 5.2, Table 5.3 and Table 5.4.

Table 5.2 shows that there is a significant improvement in the buffer times for eight out of the ten line plans. Remember that the desired minimum buffer time during peak hour on weekdays is one minute, since 30 trains pass in each direction through the central corridor. The desired minimum buffer time for line plan 3 which is a night line plan, is three minutes since only 15 trains pass in each direction through the central corridor at night. For three out of the

Line plan		Minimum buffer time in the central corridor Initial (min)	Final (min)	Impro (%)	Overall minimum buffer time Initial (min)	Final (min)	Impro (%)
1	real	0.63	1.00	+58	0.00	1.00	$+\infty$
2	real	0.73	1.00	+36	0.00	1.00	$+\infty$
3	real	3.00	3.00	+0	0.70	2.55	+264
4	random	0.33	0.64	+93	0.00	0.05	$+\infty$
5	random	0.17	0.83	+400	0.00	0.20	$+\infty$
6	random	0.37	0.99	+170	0.00	0.01	$+\infty$
7	random	0.23	0.23	+0	0.00	0.00	+0
8	random	0.23	0.23	+0	0.00	0.00	+0
9	special	1.00	1.00	+0	0.70	1.00	+43
10	special	0.92	1.00	+8	0.00	0.00	+0

Line plan		Sum of the inverse buffer times Initial (1/min)	Final (1/min)	Impro (%)
1	real	2 639	2 189	-17
2	real	2 348	2 212	-6
3	real	482	382	-21
4	random	3 293	3 323	+1
5	random	3 840	2 365	-38
6	random	3 211	2 929	-9
7	random	4 324	4 324	-0
8	random	4 357	4 348	-0
9	special	2 318	2 203	-5
10	special	3 179	3 362	+6

Table 5.2: The integrated approach significantly improves the buffer times in eight out of ten of the studied line plans.

ten line plans, the desired minimum buffer time is reached both in the central corridor and in the rest of the network. For three other line plans the desired minimum buffer time is reached in the central corridor but not in the rest of the network. Furthermore, we see that the sum of the inverse buffer times between train pairs in every station they have in common decreases, which means that the buffer times themselves increase as desired. Moreover, the results for the sum of the inverse buffer times are very similar to the minimum buffer time results in the central corridor and in the overall network. We note that a big absolute improvement of the minimum buffer time in the central corridor (or of the overall minimum buffer time) corresponds to a big improvement in the sum of the inverse buffer times, and vice versa.

Unfortunately, for two out of the ten line plans (7 and 8) no improvement in the minimum buffer time is achieved. To identify the critical line in Step 2 of the integrated algorithm, the buffer times are categorized as zero, smaller than 30 seconds, smaller than one minute and bigger than one minute. We observe that for the timetables corresponding to the initial line plans of 7 and 8 almost half of the minimum buffer times between line pairs are smaller than 30 seconds, while for the other line plans this is at most one third of the minimum buffer times. As a possible explanation, we indicate briefly that for these line plans almost every line has a pairwise minimum buffer time with some other line of below half a minute. We may therefore expect that multiple lines must be modified to see an improvement. We typically change one single line in every iteration and in case of line plans 7 and 8, it may take more than three non-improving iterations before seeing an improvement, given that every line plan we consider has between six and ten lines.

The buffer times in Table 5.2 appear to be small. However, as discussed earlier, the minimum desired buffer time in a daytime week line plan is one minute, which constitutes the value of the stopping criterion in Step 2 of the integrated approach. The upper bound on the minimum buffer time everywhere in the network is also restricted by this upper bound on the minimum buffer time in the central corridor. Moreover, even if the buffer time between two trains is zero, the timetable is still feasible. A zero buffer time means that the second train reserves a station area immediately after the first train releases this station area. Note that the release time of the first train implies that this train is already sufficiently far away. However, a zero buffer time is undesirable and any delay of the first train is immediately propagated to the second. In the case study, a line plan performs best if it allows the desired buffer time of at least one minute between every two trains. Line plan 3 is an exception for which the desired buffer time is three minutes (i.e. only 15 trains in each direction have to pass through the central corridor during night time on a weekday). The desired buffer time is achieved for line plans 1, 2 and 9 in the entire network and in line plans 3, 6 and 10 in the central corridor.

One explanation why certain line plans do not reach the desired value could be that no further improvement was made, as a result of the fact that at each iteration, the same line was identified as being critical. Either changing this line was no longer feasible or changing this line was feasible, but did not result in acceptable solutions. If changing the critical line is not feasible, then, within the same iteration, the second most critical line of the last found line plan is chosen. In the current algorithm however, if the critical line of one iteration does not give rise to good results in the next iterations, there is no backtracking to a previous iteration to try the second most critical line.

Line plan		Operator cost Initial ($\cdot 10^5$)	Operator cost Final ($\cdot 10^5$)	Operator cost Change (%)	Passenger cost Initial ($\cdot 10^7$)	Passenger cost Final ($\cdot 10^7$)	Passenger cost Change (%)
1	real	6.79	6.84	+0.74	4.17	4.23	+1.47
2	real	6.84	7.21	+5.40	4.22	4.21	-0.12
3	real	3.40	3.43	+0.64	1.05	1.06	+1.08
4	random	6.25	6.64	+6.23	4.24	4.27	+0.87
5	random	6.48	6.80	+4.94	4.27	4.29	+0.36
6	random	6.66	6.74	+1.13	4.12	4.14	+0.51
7	random	7.02	7.02	+0.00	4.09	4.09	+0.00
8	random	8.27	8.32	+0.71	4.05	4.04	-0.22
9	special	7.15	7.14	-0.17	4.43	4.44	+0.32
10	special	9.00	9.01	+0.20	4.35	4.30	-1.06

Table 5.3: For seven out of the ten line plans, the difference in operator cost and passenger cost is smaller than 1.5% when applying the integrated approach.

In Table 5.3, the operator cost and the passenger cost for the initial and final line plans are presented. We observe that for seven out of the ten line plans, the difference in operator cost and passenger cost is smaller than 1.5% when applying the integrated approach. Some plans do improve for one measure but become worse for another and although it is possible for both to improve (since we allow lines that were not in the original line pool), we do not observe this here. Note that for line plans 4 and 5 we do see a relatively large increase in operator cost (6.23% and 4.94%), combined with an increase in passenger cost, which may be a relatively large cost to pay for timetable improvement in terms of robustness. In contrast, for line plan 2, we see a similarly large increase in operator cost but a reduction in passenger cost. Here the impact must be judged by the perceived relative importance of the two measures together with the expected effect on the robustness of the service. Since it holds that for each of the line plans the difference between the passenger cost calculated based on the final timetable and the passenger cost estimated by the line planning module is less than 0.05%, we did not present these results separately.

Table 5.4 presents some characteristics of the integrated approach. We indicate under which stopping criterion the algorithm was terminated. For three out of the ten line plans the desired minimum buffer time is achieved in the central corridor and in the rest of the network ('DES') and for the remaining seven the algorithm ended with the best found value ('BFV'). Table 5.4 also reports how many iterations the integrated approach passed through before a stopping criterion was reached. This value ranges between one and seven. We report the number of out-of-pool lines which are in the final solution, referring to lines

Line plan		Stopping criterion	# Iter	# Out-of-pool lines	Average run time timetabling (min)
1	real	DES	4	5	183.40
2	real	DES	3	5	4.88
3	real	BFV	2	3	0.50
4	random	BFV	6	5	75.71
5	random	BFV	7	7	385.56
6	random	BFV	7	5	167.19
7	random	BFV	3	0	126.75
8	random	BFV	3	1	9.25
9	special	DES	1	2	47.50
10	special	BFV	5	3	346.83

Table 5.4: Characteristics of the integrated approach.

that are in the final solution but do not come from the original, restricted, line pool. These out-of-pool lines are similar to a line in the pool, but have a modified stopping pattern. We observe that the five line plans with the highest number of out-of-pool lines (line plans 1, 2, 4, 5 and 6) have the largest relative improvement of the minimum buffer time in the central corridor; have the largest increase in operator cost; and (with one exception) have the highest increase in passenger cost. Therefore, while including new lines in the line pool has the potential to improve the minimum buffer times significantly, it may have a negative effect on passenger and operator costs.

The final characteristic in Table 5.4 is the run time for timetabling. The total run time for timetabling consists of the creation of the optimal timetable in Step 2 of the integrated approach for each iteration and of the determination of the initial timetable. As described in Section 5.4, the timetabling model is solved sequentially with two objective functions at each iteration. Firstly, the buffer time in the central corridor is maximized, and secondly the buffer time in the rest of the network is increased with a bound on the buffer time in the central corridor fixed by the first step.

If the algorithm for example stops after three iterations for line plan 2, this means that eight timetables are calculated: two initially (for the two optimization criteria) and two at each of the three iteration steps. The average run time for timetabling while optimizing line plan 2 is 4.88 minutes, which means that the run time for each calculated timetable in the integrated approach is on average 4.88 minutes. All timetables are calculated with CPLEX 12.6 on an Intel Core i7-5600U CPU @ 2.60 GHz. We observe that there is a high variability in the average run times for the different line plans. Moreover, a high computation time may occur in case a big improvement is observed (line plans 1 and 5) and also when either no improvement (line plan 7) or only a small improvement (line plan 10) is observed. Furthermore, the run time for timetabling can differ significantly from one iteration to the next. Even if two line plans are not dissimilar, one can be intrinsically more difficult to solve. An explanation could be that due to changes in the stopping pattern, trains of different lines are more or less susceptible to catching up with each other in 'the fingers' of the S-tog network, resulting in trains being more complex to be spread optimally. The timetabling module runs to optimality (relative gap smaller than 0.05%) for about 85% of the timetables. The average run time per timetable optimized within the time limit of 12 hours is 3801 seconds. For the other optimizations a time limit of 12 hours is imposed. The line planning module for the selected line pool determined by the critical line, runs to optimality in all instances, taking at most up to ten minutes for cases where many lines have to be changed.

Finally, from Tables 5.2, 5.3 and 5.4 can be deduced that line plans 5 and 9 did not converge to the same final line plan and line plans 8 and 10 did not either. We see that the final line plan and timetable for line plans 9 and 10 score better in terms of passenger robustness, i.e. the minimum buffer time between line pairs is larger, while line plans 5 and 8 score better in terms of operator and passenger cost. Based on these results, the final decision on which a line plan is preferred, rests with the operator. In our opinion, the optimized version of line plan 1 will be most passenger robust taking into account the corresponding expense for the operator and the passengers.

5.5.2 Illustration

In order to illustrate the integrated approach, we apply it to line plan 2. As this is an existing line plan, we skip Part 1 of the algorithm and only look at the iterative steps in Part 2. The estimated operator cost of this line plan is $6.84 \cdot 10^5$, the estimated total passenger travel time is $4.22 \cdot 10^7$. The optimal value for the minimum buffer time for this line plan in the central corridor of the network is 0.73 minutes. The optimal value for the overall minimum buffer time if the minimum buffer time in the central corridor is bounded below by 0.73 minutes, is zero minutes. The minimum buffer times between the line pairs are shown in minutes in Table 5.5. The smallest buffer time between line i and j is the same as the smallest buffer time between line j and i, so Table 5.5 is in fact symmetric, but the superfluous information is omitted here. If two lines do not share a part of the network, the minimum buffer time between these lines is indicated as 60 minutes, which is the period length of the cyclic timetable.

The smallest buffer time in Table 5.5 is zero minutes. This buffer time is

Line	0	1	2	3	4	5	6	7
0	1.73	0.73	0.73	2.47	2.88	3.13	2.47	60
1	-	**0.00**	2.47	0.73	1.15	2.57	0.73	60
2	-	-	2.2	1.30	1.47	1.30	3.50	60
3	-	-	-	7.17	0.88	1.40	6.70	60
4	-	-	-	-	4.45	7.58	1.03	60
5	-	-	-	-	-	2.87	0.80	60
6	-	-	-	-	-	-	6.97	60
7	-	-	-	-	-	-	-	0.57

Table 5.5: First iteration: The overall minimum buffer time is zero minutes, if the minimum buffer time in the central corridor is bounded below by 0.73 minutes.

between line 1 and itself. This means that the turn platform of line 1 in one of its terminal stations is permanently occupied by a train of the first line. Obviously, the critical line here is line 1.

The line planning module adds a stop to line 1 by considering only the line pool that contains alternatives for line 1 of the same frequency in the second iteration. The new operator cost is estimated to increase to $6.99 \cdot 10^5$ and the new total passenger travel time to slightly increase to $4.23 \cdot 10^7$. The optimal value for the minimum buffer time for this line plan in the central corridor of the network is 1.00 minute. The optimal value for the overall minimum buffer time, if the minimum buffer time in the central corridor is bounded below by 1.00 minute is still zero minutes. The minimum buffer times between the line pairs of the first modification of line plan 2 are present in minutes in Table 5.6.

Line	0	1	2	3	4	5	6	7
0	6.17	1.01	2.99	60	3.01	1.00	3.99	2.99
1	-	2.14	2.99	60	1.00	1.99	1.00	1.00
2	-	-	0.29	60	1.00	1.00	1.01	1.00
3	-	-	-	0.67	60	60	60	60
4	-	-	-	-	1.82	2.99	2.99	2.99
5	-	-	-	-	-	7.23	6.99	6.99
6	-	-	-	-	-	-	**0.00**	1.00
7	-	-	-	-	-	-	-	3.02

Table 5.6: Second iteration: The overall minimum buffer time is zero minutes, if the minimum buffer time in the central corridor is bounded below by 1.00 minute.

The smallest buffer time between two lines is still zero minutes. This buffer time is now only associated with line 6. Again, this means that the turn platform of line 6 in one of its terminal stations is permanently occupied by a train of line 6. The critical line is line 6.

In Step 2 of the third iteration, the line planning module first considers line pools that only contain alternatives for line 6 that have the same frequency, but these do not lead to a feasible line plan. We then consider the line pool that contains alternative lines for line 6 for different frequencies. The result is a feasible line plan that does not include original line 6 and 7, each of frequency three, but contains a new line of frequency six. The original line 6 stops at the same stations as the original line 7, but has some additional stops at one end of the line. The new line is the combination of the original lines 6 and 7. The new operator cost is estimated to mount to $7.22 \cdot 10^5$ and the new total passenger travel time to $4.20 \cdot 10^7$. The optimal value for the minimum buffer time of this line plan in the central corridor of the network remains 1.00 minute. However, the optimal value for the overall minimum buffer time, if the minimum buffer time in the central corridor is bounded below by 1.00 minute, has now increased to 0.70 minutes. The minimum buffer times between the line pairs in the third iteration are present in minutes in Table 5.7. The smallest buffer time is now only associated with line 3, so the new critical line is line 3.

Line	0	1	2	3	4	5	6
0	10.16	2.99	1.00	60	2.99	8.98	1.00
1	-	2.10	2.99	60	1.00	2.01	1.00
2	-	-	1.15	60	1.00	1.01	3.01
3	-	-	-	**0.70**	60	60	60
4	-	-	-	-	2.06	3.01	2.99
5	-	-	-	-	-	8.05	1.00
6	-	-	-	-	-	-	1.69

Table 5.7: Third iteration: The overall minimum buffer time is 0.70 minutes, if the minimum buffer time in the central corridor is bounded below by 1.00 minute.

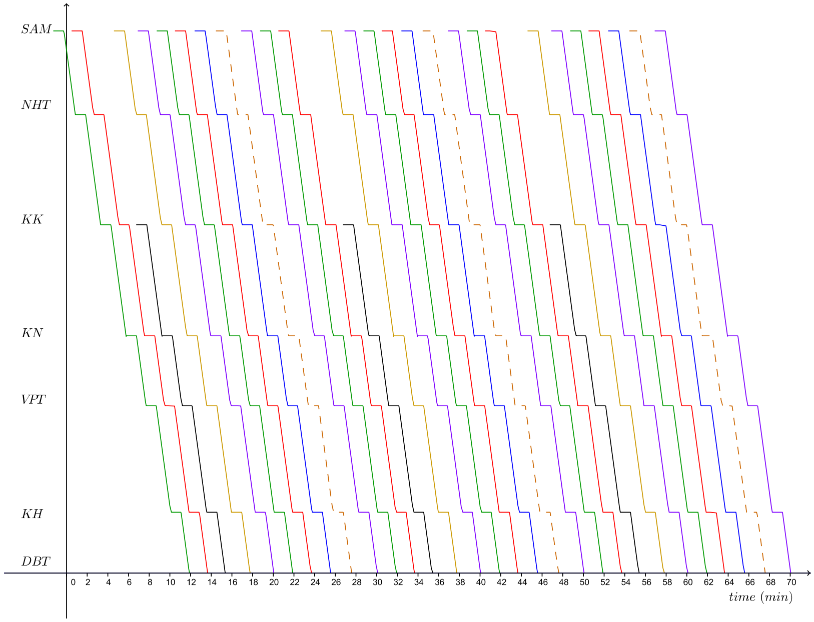
The line planning module skips a stop of line 3 in the fourth iteration. The new operator cost is estimated to mount to $7.21 \cdot 10^5$ and the new total passenger travel time to $4.21 \cdot 10^7$. The optimal value for the minimum buffer time of this line plan in the central corridor of the network is still 1.00 minute. The optimal value for the overall minimum buffer time if the minimum buffer time in the central corridor is bounded below by 1.00 minute is now 1.00 minute.

The minimum buffer times between the line pairs in the fourth iteration are

Line	0	1	2	3	4	5	6
0	10.20	3.01	1.00	60	2.99	8.98	1.00
1	-	2.15	2.99	60	1.00	1.99	1.00
2	-	-	1.09	60	1.00	1.00	2.99
3	-	-	-	1.00	60	60	60
4	-	-	-	-	2.11	2.99	2.99
5	-	-	-	-	-	8.02	1.00
6	-	-	-	-	-	-	1.66

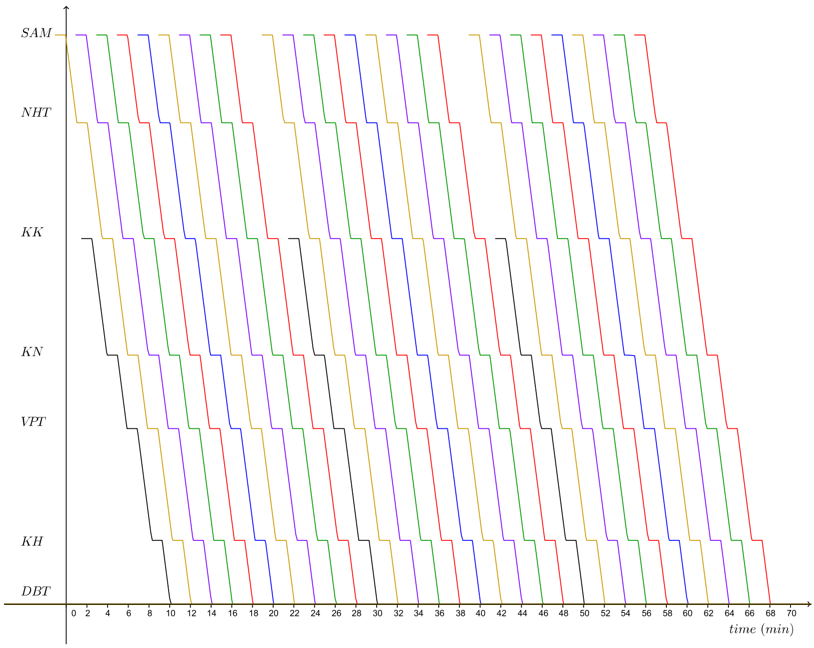
Table 5.8: Fourth iteration: The overall minimum buffer time is 1.00 minute, if the minimum buffer time in the central corridor is bounded below by 1.00 minute.

shown in minutes in Table 5.8. This overall minimum buffer time is closer than five percent to the minimum desired buffer time of one minute, so this is the last iteration of the algorithm.



(a) Time-distance diagram for the initial timetable of line plan 2.

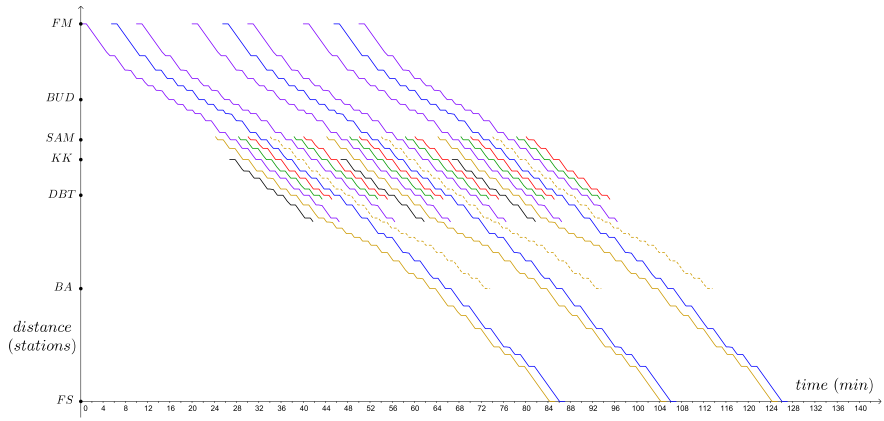
Figure 5.8 and Figure 5.9 present the time-distance diagrams of the initial and the finally selected timetable for line plan 2, corresponding to Tables 5.5



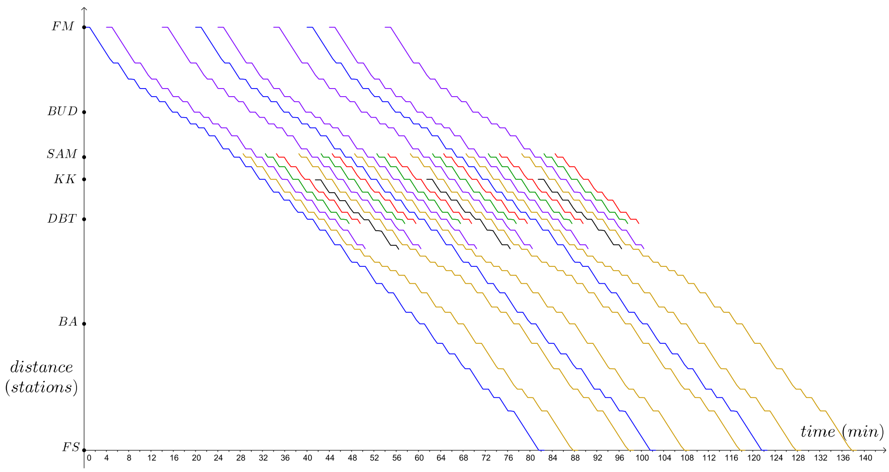
(b) Time-distance diagram for the finally selected timetable of line plan 2.

Figure 5.8: These time-distance diagrams visualize the timetables restricted to the central corridor for the trains from north to south (Svanemøllen (SAM), Nordhavn (NHT), Østerport (KK), Nørreport (KN), Vesterport (VPT), København H (KH), Dybbølsbro (DBT)). The colors indicate the different lines. The same color is used to indicate lines between the same terminal stations in both diagrams.

and 5.8 respectively. Figure 5.8 zooms in on the timetable for the central corridor, whilst Figure 5.9 presents the timetable of the central corridor and two fingers of the network. These time-distance diagrams only show the trains driving from north to south. Horizontal parts of a time-distance path represent station activities in station areas where the train stops. The other parts of a time-distance path include running activities between two station areas and station activities in station areas where the train does not stop. Both the initial and the finally selected timetable have 30 trains per hour passing through the central corridor. However, the initial timetable has seven lines passing through the central corridor, while the finally selected timetable has only six different lines passing through the central corridor.



(a) Time-distance diagram for the initial timetable of line plan 2.



(b) Time-distance diagram for the finally selected timetable of line plan 2.

Figure 5.9: These time-distance diagrams visualize the timetables of the finger from Farum (FM) to Svanemøllen (SAM), the central corridor (Svanemøllen (SAM), Nordhavn (NHT), Østerport (KK), Nørreport (KN), Vesterport (VPT), København H (KH), Dybbølsbro (DBT)) and the finger from Dybbølsbro (DBT) to Frederikssund (FS) for the trains from north to south. The colors indicate the different lines. The same color is used to indicate lines between the same terminal stations in both diagrams.

In Figures 5.8 and 5.9, the yellow-brown solid line and the yellow-brown dashed line of frequency three are combined to one line of frequency six during the line planning phase of the third iteration, as discussed in the illustration above. For both timetables there is one line that uses the circle track, so this line does not pass through the central corridor. Both timetables also have one line which route starts inside the central corridor. In Figure 5.8, we can observe that the trains are regularly spread in the finally selected timetable, while this is not the case for the initial timetable. Furthermore, in Figure 5.9, we see that the final timetable spreads the lines better in the lower finger and equally well in the upper finger. Note that the yellow-brown line has more stops than the blue line in the lower finger, since its time-distance paths contain more horizontal parts. By combining the yellow-brown solid and dashed line, it became possible to put the slower yellow-brown line after the faster blue line when leaving the central corridor. This avoids the blue line catching up with the yellow-brown line as is the case in the initial timetable.

5.5.3 Limited shunt capacity

Here the impact of limited shunt capacity in terminal stations on the performance of the railway service is quantified. If trains have to turn on their platform in a terminal station, they occupy the platform for several minutes. A more preferred option, from a timetabling point of view, would be that a train disappears from the platform in its terminal station after dwelling and only appears again when departing for a subsequent trip. Moreover, in the shunt, there is always a train ready to depart for the next trip leaving from that terminal station, for the case that the previously arrived train is not ready yet. In this way, the timetable for a line and its opposite line can be detached in the terminal stations. Furthermore, the train will not interfere with other trains that dwell on the platforms in the terminal station. However, this option is only possible if the train can stay in a flexible and large enough shunt. A schematic view of a terminal station without and with a flexible shunt is provided in Figure 5.10

For each of our ten initial line plans, we compare the minimum buffer times of two timetables. The two timetables are constructed taking into account that trains either have to turn on their platform or make use of a flexible shunt. These timetables are calculated up to optimality or until the time bound of 12 hours is reached. Timetables that take flexible shunts into account can be obtained by using model (5.31)-(5.33) and omitting the turn activities, the buffer activities on platforms (in terminal stations) and the turn connection activities in constraints (5.33) of the timetabling model. The impact on the performance is evaluated by the minimum buffer time in the central corridor

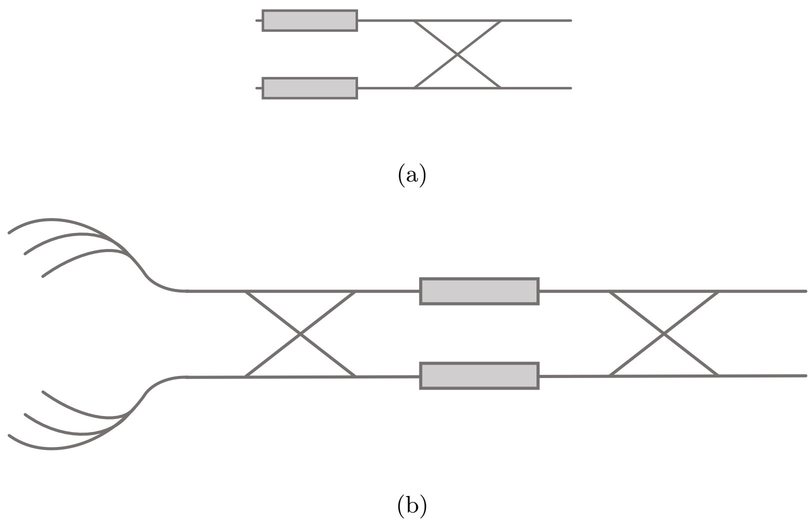


Figure 5.10: Part (a) shows a terminal station without shunting space. Part (b) shows a terminal station with a flexible shunt. In the latter the trains can use several tracks behind the platforms where they can wait and turn for their next trip. There is also a junction available behind the platforms to switch platforms.

and the overall minimum buffer time between trains while keeping the best found value for the central corridor.

The results can be found in Table 5.9. The column label *'No shunt'* refers to the timetable where trains have to turn on their platform, the label *'Flex shunt'* refers to the timetable where trains can turn in a flexible shunt. The timetables are constructed with CPLEX 12.6 on an Intel Core i7-5600U CPU @ 2.60 GHz. Timetables that are not calculated up to optimality in the provided 12 hours are indicated with *'*'*. The average calculation time for the timetables that take a flexible shunt into account is 6624 seconds, while the average calculation time for the timetables where trains have to turn on their platform equals 3053 seconds. Only calculation times are taken into account in case optimality is achieved (and proven). The much higher average calculation time for the case with flexible shunts can be explained by the fact that there are many more feasible solutions for this case. The set of feasible solutions for the case without flexible shunt is a subset of the set of feasible solutions for the case with flexible shunt.

We see that the minimal overall buffer time and the minimum buffer time in the central corridor equal the desired value in eight out of ten line plans in case the terminal stations are equipped with a flexible shunt. In case trains have to turn on their platform, only two of the timetables allow this desired value in the central corridor. Also for the overall minimum buffer time, while restricted by the minimum buffer time in the central corridor, much better values can be obtained. Here six out of ten line plans equal the desired value in case the terminal stations are equipped with a flexible shunt, instead of none if trains have to turn on their platform. Note that optimal values of both cases provide railway companies with the information on how much the performance could be increased if large enough flexible shunts would be built in terminal stations.

		Minimum buffer time in central corridor (min)		Overall minimum buffer time (min)	
Line plan		No shunt	Flex shunt	No shunt	Flex shunt
1	real	0.63	1.00	0.00	1.00
2	real	0.73	1.00	0.00	1.00
3	real	3.00	3.00	0.70	3.00
4	random	0.33	1.00	0.00*	1.00
5	random	0.17*	1.00	0.00	0.00*
6	random	0.37	1.00	0.00	0.00
7	random	0.23	1.00	0.00	0.00
8	random	0.23	0.60*	0.00	0.00
9	special	1.00	1.00	0.70	1.00
10	special	0.92	1.00	0.00	0.50

Table 5.9: The minimal buffer times improve if the infrastructure allows to turn in a shunt.

We conclude that for different kinds of line plans, the probability of propagation of delays can be highly reduced if the terminal stations have a flexible shunt, so trains don't need to turn on (and occupy) their platform during the complete turn movement. By either omitting or including the constraints concerning turning in the terminal stations (constraints of the turn activities, the buffer activities on platforms in terminal stations and the turn connection activities), we have calculated the actual value of this potential improvement for the DSB S-tog network of Copenhagen. This potential improvement is the difference in minimum buffer time of the *'Flex shunt'* columns compared to the *'No shunt'* columns in Table 5.9.

5.6 Conclusion

This chapter presents a heuristic algorithm that builds a line plan from scratch, resulting in a feasible and passenger robust timetable. The method iterates interactively, alternating between a line planning module and a timetabling module, improving the robustness of an initially built line plan. Both modules consist of an exact optimization model. The line planning module optimizes a weighted sum of passenger and operator costs, while the timetabling module focuses on improving minimum buffer times between line pairs. Appropriate and sufficiently large buffer times between train pairs are needed to reduce the risk of delays being propagated from one train to the next, thereby obtaining a passenger robust railway schedule. The timetabling module identifies a critical line based on the minimum buffer times between line pairs. The line planning module creates a new line plan in which the time length of the critical line is changed. Changing the time length of a line may create more flexibility in the schedule, which may result in improvements in passenger robustness. The approach was tested for ten different line plans on the DSB S-tog network in Copenhagen. This is a high-frequency railway network with 84 stations, currently nine lines and restricted shunt capacity in the terminal stations. For eight out of ten initial line plans the passenger robustness could be significantly improved, while the changes to the line plan in most cases resulted in a change of less than 1.5% in the weighted sum of operator and passenger cost. Ultimately the operator needs to make the final decision based on the preferred criterion, considering the measures presented here or others we have not captured. This chapter also shows the potential of constructing passenger robust timetables in case the railway network is equipped with flexible shunt areas in its terminal stations.

5.7 Future research

An initial idea for future research is a smart extension of the integrated approach to overcome the situation where a certain line remains critical in each iteration, while keeping the computation time restricted. Another extension could be to allow different shunt characteristics in different terminal stations. In the presented research we needed to cope with the actual very strict requirement present in the DSB S-tog system reality requiring to have a schedule in which no train uses shunt capacity in a terminal station during daily operation. Furthermore, the development of a single integrated exact model that combines line planning and robust timetabling which is solvable in a reasonable amount of time for other real networks (similar to the DSB S-tog network) would be a next noteworthy step. A first idea to solve this model is a decomposition in a timetabling part and a line planning part where the coupling constraints assure the compatibility of the solutions to both problems. An alternative therefore is to replace the timetabling model with a faster timetabling heuristic, which focuses on the same objectives. A heuristic for the DSB S-tog is constructed and starts with detecting a line order for which the desired spreading in the central corridor can be achieved (Pieters, 2017). If such a line order cannot be found, concessions are made on the desired buffer time in the central corridor. Further investigations to include the restricted shunt capacity in this heuristic are still necessary.

A further idea for future research is to remove the requirement that trains of a line must operate exactly evenly timed (e.g. once every ten minutes for a six-per-hour line). This requirement is consistent with the way the network is currently operated and ensures a regular service for customers. However it is potentially severely restrictive for the timetable given the tight spacing of trains in the central corridor. Relaxing this requirement could increase the complexity of the timetabling model, both by expanding the solution space and by requiring new constraints and possibly an objective measure for the evenness of train timings.

5.8 Acknowledgments

This research is partly funded by a travel grant for a long stay abroad financed by Research Fund - Flanders (FWO). Together with my supervisor, I thank Richard M. Lusby and Simon H. Bull for the nice collaboration on this research project. I would also like to thank the Danish railway operator DSB for their help and commitment during the entire project.

6 Conclusion and future research

This chapter concludes the dissertation, brings together the ideas for future research and summarizes how this research can be implemented in practice in cooperation with the railway companies.

6.1 Conclusion

The main goal of public transport is to transport passengers. Therefore, it seems logical to take passengers' preferences into account while planning public transport. Striving for passenger robustness embraces short travel times and reliable travel times which are two of the top priorities of public transport users. Precisely for this reason we advocate a passenger robust schedule. This is a schedule that minimizes the total passenger travel time in practice, which means in the context of frequently occurring small delays. Striving for passenger robustness while making railway schedules is the common thread linking the chapters in this dissertation.

When developing a new timetable for the Belgian railway network, it is still current practice to introduce large chunks by hand and to carry out manual try-outs. This is a cumbersome, time-consuming and suboptimal process that, as has been shown in this dissertation, can be improved upon a great deal. The current planners compare the performance of a newly constructed timetable with the performance of the current timetable through simulation. However, no absolute performance or optimization criterion is premised. This working method can be acceptable if many new timetables are constructed and compared to each other in order to increase the probability of finding a really good timetable. However, the need for quality is counteracted by the time it costs to achieve these goals. A first solution proposed in literature before, is to improve an existing timetable and this based on an absolute criterion (Dewilde, 2014b). Therefore, passenger robustness is premised. This evolution makes it possible to reduce the number of timetables that must be constructed from scratch to a few, to assure a certain performance level. In so-doing, it highly improves the performance to time ratio. In this dissertation in Chapter 3, the state of the art on this topic is further improved by taking passenger numbers and recurring delays explicitly into account during the optimization. In that way, the passenger robustness for the Brussels station area, the main bottleneck in the Belgian railway system, could be improved by 4% compared to the state of the art and by 11% compared to a reference timetable from the Belgian railway infrastructure manager Infrabel.

A clear alternative to improve the performance to time ratio is to make a passenger robust (cyclic) timetable from scratch. Then, only one timetable has to be constructed and no post-processing phase is necessary. Moreover, the final timetable does not depend on the initial, possibly badly performing, timetable. A timetabling and routing model that construct passenger robust timetables and routing plans from scratch is a second achievement in this dissertation, as was discussed in Chapter 4. These models are especially useful to make a schedule in the context of railway bottlenecks with complex infrastructure and dense train traffic. They exploit the advantages of adding extra spreading between trains (buffer times). Combined with simulation, these models are also adding extra time to the train travel times themselves (supplements) in an automated way. This methodology improved the passenger robustness for Brussels by 8% compared to the state-of-the-art timetable in literature and by 17% compared to the reference timetable from Infrabel. An extra advantage of the routing model is that it results in the routing plan with the least interactions between the trains, without needless detours in the routes. This routing plan is ideal as a reference to assess the quality of other routing plans. Extra features of the timetabling model are that it includes constraints to take passenger transfers and the splitting, coupling and re-usage of trains into account.

Another problem that railway companies are facing is due to the fact that railway planning is done in several phases. This often causes that a result of a previous planning phase does not assure a (good) result for the next phase. A third contribution of this dissertation, described in Chapter 5, is a method that builds a line plan and a timetable together. This method assures normative macroscopic feasibility, passenger robustness and also focuses on minimizing operator costs. However in the end (in order to make a final decision), the railway company can decide on the weights of the different objectives. This dissertation also presents and proves some theoretical results and insights on how infeasible lines, infeasible line combinations or infeasible line plans can be identified. Implementation of these theoretical results shows that the limited shunt capacity available in the terminal stations of the DSB S-tog network in Copenhagen largely affect the potential to build feasible and passenger robust line plans and timetables.

My PhD taught me that railway research is a very challenging research area, in which improvements could be directly applied in a real world context. In the last 20 years, lots of breakthroughs have been achieved in railway research. However, methodologies and conclusions are often network dependent, i.e. the functionality and efficiency is determined by the complexity of the infrastructure and the density of the railway traffic. The Belgian network is particularly complex due to the railway bottleneck in Brussels. This bottleneck is characterized by a small tunnel, 'impossible' to widen, connecting Belgium's three busiest stations. Concerning this Belgian railway network, methodologies have already been published for which is shown that they are able to improve the passenger robustness of an existing microscopic timetable in Brussels and methodologies that make a passenger robust macroscopic timetable and do platform allocations for the rest of the Belgian country, (Vansteenwegen, 2008, Dewilde, 2014b, Sels, 2016b). The developed methodologies don't need manual calculations and are much faster to solve than the current practice in Belgium. Furthermore, research is going on on how to deal with real-time conflicts and the results are promising. This dissertation completes the missing part in the state of the art by adding a methodology to construct a passenger robust microscopically conflict-free timetable for a railway bottleneck, *from scratch*, without manual calculations. Moreover, this dissertation shows how to coordinate timetabling and railway routing on microscopic level and line planning and timetabling on macroscopic level and how to deal with limited shunt capacity or turn restrictions in terminal stations. Most of these very promising results are published in literature and even more information is available for the use of the Belgian railway companies. I believe that it is about time that the Belgian railway companies bring this state of the art to practice if they want to improve the service they offer to the passengers.

This conclusion ends with a list of the main contributions, organized per chapter.

Main contributions of Chapter 3 are:

- A method to use passenger data while improving the passenger robustness of an existing timetable.
- A method to use historical delay data while improving the passenger robustness of an existing timetable.

- A validation of both approaches separately and of combinations of both approaches on a real, large and complex railway bottleneck, namely Brussels (Belgium).
- Resulting routing plans and timetables for Brussels, which are more passenger robust than reference routing plans and timetables from practice and from the literature.

Main contributions of Chapter 4 are:

- A routing model and a timetabling model to construct a conflict-free and passenger robust microscopic routing plan and timetable from scratch.
- An iterative approach including the routing and the timetabling model and a feedback loop that helps constructing even more passenger robust schedules by considering the inclusion of supplements as an extra degree of freedom.
- A validation of the routing model, the timetabling model and the iterative approach on a realistic, large and complex railway bottleneck, namely Brussels (Belgium).
- Passenger robust routing plans and timetables, built from scratch, for Brussels, which significantly improve the passenger robustness of reference routing plans and timetables from practice and from literature.
- Additive and alternative constraints to speed up routing and timetabling models and to include transfers, re-usage, splitting and coupling of trains into these models.

Main contributions of Chapter 5 are:

- The integration of line planning, timetabling and passenger robustness.
- An approach that builds coordinated line plans and timetables *from scratch.*
- Two insights and proofs on timetable-infeasibility of line plans.
- The inclusion of limited shunt capacity of terminal stations in line plan and timetable optimization.
- Practical conclusions for the DSB S-tog network in Copenhagen based on experimental results.

6.2 Future research

This section brings together the ideas for future research that are presented throughout this dissertation.

6.2.1 More degrees of freedom in timetabling

A first thought for future research is to take more degrees of freedom into account during timetable optimization. We elaborate on five examples.

Example 1 This dissertation presents a method to account for recurring *arrival* delays in railway timetabling and routing. This method could be extended by also accounting for recurring *dwell* delays. However, it is more complex to subtract information on dwell delays from historical delay data. This is because dwell delays occur inside the considered network, so it might be that the train already suffers from delays earlier on. The difficulty is to distinguish which part of the train delay consists of earlier delays or is due to them and which part is related to the actual new dwell delays.

Example 2 This dissertation presents an iterative railway routing and timetabling approach that optimizes buffer times and supplements. A disadvantage of this method is that only one supplement (at one location) can be added per train per iteration. An idea might be to make decisions on multiple supplements for one train (at different locations) at the same time. Adding multiple well-chosen supplements might speed up the iterative approach. It then might be necessary, however, to provide an upper bound on the amount of supplements that can be added in each iteration or per train to avoid an overgrowth of supplements. The more supplements are added, the higher the capacity usage will be and the more difficult it will become to construct a conflict-free timetable. A second idea concerning the automatic inclusion of supplements is to make the assignment dependent on the number of passenger travel times that will be affected. Now, this assignment is based on knock-on delays and non-used supplements, which are two performance indicators based on train level and not on passenger level. A third idea concerning the automatic inclusion of supplements is to include it into the exact timetabling model, which means transforming the iterative approach into an integrated model. With the current objective to maximize the buffer times, the model would generally not include running and dwell time supplements in order to keep the travel times as short as possible. Supplements would only be added to increase buffer times somewhere further in the network and not necessarily where they are useful to absorb delays. So an appropriate objective function must be chosen.

Example 3 The timetabling methods presented in this dissertation are theoretically able to account for passenger transfers. However, no data is available to investigate the effect of passenger transfers in Brussels on the passenger robustness. It would be very interesting to analyze the actual effect of passenger transfers on the schedule and the passenger robustness. To perform these tests, a timetable that accounts for passenger transfers and a timetable that does not account for passenger transfers can be compared by simulating both timetables while using the passenger transfer data. The performance of the timetable that accounts for passenger transfers is dependent on the amount of transferring passengers. If the timetable accounts for transfers in which no passengers are interested, the extra constraints to assure these transfers will negatively affect the performance of the timetable. If a transfer is useful for a large amount of people, one can expect that taking this transfer into account will improve the performance of the timetable. In the ideal case, the effect of taking passenger transfers into account is measured by performing several tests in which the amount of transferring passengers is varied. These tests might also give interesting insides for railway companies on which transfers to take into account and which not.

Example 4 This dissertation shows that the shunt characteristics in terminal stations have a large impact on the construction of a passenger robust timetable. It is therefore important to model the shunt characteristics at stations as realistically as possible.

Example 5 Currently, the strict requirement that trains of a line must operate exactly evenly timed (e.g. once every ten minutes for a six-per-hour line) is consistent with current operational practice in the Copenhagen DSB S-tog network and ensures a regular service for customers. However, this requirement is potentially severely restrictive for the timetable, given the tight spacing of trains in the central corridor of the DSB S-tog network. An idea is to abandon this strict requirement. However, relaxing this requirement could increase the complexity of the timetable model, both by expanding the solution space and by requiring new constraints and possibly an objective measure for the evenness of train timings.

6.2.2 Fast heuristic for timetabling

Including more degrees of freedom in the timetabling methods presented in this dissertation might lead to a better overall solution, but will make the problem harder to solve and will thus lead to longer computation times. An exact model with a passenger travel time objective that is solved in a reasonable amount of time for microscopic timetabling in complex railway station areas is a big challenge. This leads us to another thought for future research: developing a fast timetabling (meta)heuristic that can deal with bottlenecks as is the case in Brussels and which can be easily updated to take one or more of the above extensions into account. This heuristic, in the ideal case with a stochastic passenger travel time objective for microscopic timetabling, could then weigh the advantages of both buffer times and supplements in the construction of a timetable. When considering the Brussels network, it seems worthwhile to start with railway routing and thereafter making an assignment of timings in Brussels-Central. Trains using the same platform in Brussels-Central could be equally distributed over the period of the cyclic timetable. Thereafter it could be checked whether there are conflicts in the grid zones (and later in Brussels-South, Brussels-North, Brussels-Schaarbeek and on the outgoing tracks of the network). In case of a conflict, either the order of trains could be changed, or depending on the blocking time overlap, trains could be shifted in time. An algorithm based on these principles is constructed for the DSB S-tog network in Copenhagen. For Brussels, however, there is no proof of concept yet.

6.2.3 Global optimum for line planning and timetabling

Another thought for future research is related to the integration of timetabling and line planning. In this dissertation, a heuristic is developed that integrates an exact line planning and timetabling model. Based on the proposals for future research above, a first idea is to replace the timetabling model by a fast timetabling heuristic. Furthermore, developing a single integrated exact model that combines line planning and robust timetabling, which is solvable in a reasonable amount of time for real networks (e.g. similar to the DSB S-tog network) would be a huge achievement. The advantage would be that a global optimum can be found. Unfortunately, this problem is even harder than developing an exact timetabling model taking all features into account, as it also includes the line planning phase as a large and complex extra degree of freedom.

6.3 Future work to bring this research into practice

In order to implement the method whereby the design of a passenger robust timetable and routing plan as presented in this dissertation, into practice, the following steps must be taken.

Accurate data on passenger numbers, on demand, on passenger transfers, on the different components of blocking times and on delays must be collected. Furthermore, up-to-date microscopic infrastructure data must be available for the considered network. In order to use the routing model, a list of all possible routes between each pair of adjacent station areas must be created. This is a one-time task, assuming no changes in the infrastructure. To create this list, efficient algorithms exist in literature. Moreover, such a list is already created for the Brussels station area.

Using this information, we can create a timetable and routing plan, as we already did for the Brussels station area. The task of the railway company is then to test this newly created planning with their own simulation software. A second option would require that the railway company implements the models presented in this dissertation in their own planning system. This has the advantage that they could take additional constraints into account. The routing and the timetabling model, discussed in Chapter 4, are especially developed to give optimal results for complex, dense networks. The benefit of the presented models will be larger in this kind of network. The railway company needs to compare the developed method with other methods with respect to speed and the solution quality for networks or network parts that are less dense and complex. For the Belgian case, the routing and the timetabling model might be accompanied with the approach of Sels (2016b) to expand the timetable to the rest of the Belgian network with a view to plan trains that do not pass Brussels. Also for this second option, the railway company needs to test the newly created planning with their simulation software.

In order to implement the method whereby the design of a passenger robust line plan and timetable presented in this dissertation, into practice, the following steps must be taken.

Also here, accurate data on actual passenger numbers, on demand, on passenger transfers, on safety and service requirements, on turn restrictions, on the different components of blocking times and on delays must be collected. With the approach presented in this dissertation, we then create passenger robust line plans and macroscopic timetables. The railway company needs to define the weights that they want to assign to operator cost and passenger robustness, in order to decide upon which combination of timetable and line plan they prefer. Thereafter, the railway company can simulate the chosen line plan and timetable with its own simulation software. They especially have to check whether the created timetable is microscopically conflict-free. A second option would again require that the railway company implements the approach in its own planning system and immediately takes the microscopic infrastructure into account.

Another task for railway companies would consist of enlarging the pool of lines from which a line plan must be chosen. In Chapter 5, we observed that enlarging the pool can have a large effect on the passenger robustness, but also on the operator cost. Combined with a smart enlargement of this pool, railway companies might want to change the passengers' mind set about different stopping patterns and transfer options.

A
Appendix

A.1 Deadlocks

In this appendix, the effect of how deadlocks are handled in the simulation, is investigated. We only look at two strategies. The first strategy is to ignore the simulation runs in which a deadlock occurs. This strategy is used by Dewilde (2014b). We initially did not focus on the simulation itself and started by simply using the same simulation as in Dewilde (2014b). Therefore, we also used this same simulation strategy in all our published papers. Since we prefer that the results in this dissertation correspond to the results in the papers, the same simulation strategy was also used in all results presented in this dissertation.

Another strategy is to avoid the deadlocks by taking appropriate dispatching measures. This changes the implementation of the simulation and might lead to different experimental results obtained with this second simulation strategy.

First, the two strategies are explained in more detail in Section A.1.1. Thereafter, the results obtained by the two strategies are presented and discussed in Section A.1.2. Then a conclusion is formulated in Section A.1.3 together with some ideas for future work.

A.1.1 Strategies

The first strategy ignores simulation runs in which a deadlock occurs. This strategy works as follows. In order to get results for 10 000 simulation runs, actually 10 000 simulation runs plus the number of simulation runs in which a deadlock occurs, are executed. No data is kept, nor is analysis done on the simulation runs in which a deadlock occurs. If a deadlock occurs, the simulation run is immediately terminated and a new delay drawing is done to do this simulation run over. This is a simple strategy requiring no complex dispatching measures. Obviously, in practice, the deadlock would be prevented, but not in the simulation so we cannot measure the effect of preventing deadlocks.

The second strategy avoids deadlocks before they actually happen. In fact, the simulation runs until a deadlock occurs. Then, first, the trains that cause the deadlock and the sections that they have in common are determined. Thereafter, the time instants at which the trains entered the shared sections in this simulation run are compared. The train that entered the shared infrastructure the first, gets priority. The train that entered the shared infrastructure the latest will have to wait until the other train leaves the shared infrastructure. The time and all variables are restored to the time instant just before this last train entered the common section. This last train gets a knock-on delay from the first train. This knock-on delay equals the time that the first train needs to leave the shared infrastructure, starting from the time instant to which the time is restored. In fact, this strategy corresponds to a first come first serve strategy and is a straightforward strategy to avoid deadlocks.

A.1.2 Results

This section presents and compares the results obtained by both simulation strategies. The performance indicators presented in the tables below coincide with the performance indicators in Chapter 4. The columns are labeled *'Remove'* and *'Avoid'*, which refer, respectively, to the first strategy, where simulation runs with a deadlock are *removed*, and the second strategy, where deadlocks are *avoided*. Table A.1 shows the passenger robustness for both strategies together with the percentage improvement in passenger robustness compared to the reference planning of the Belgian railway infrastructure manager Infrabel. The results of each strategy are compared to the evaluation of Infrabel's timetable with this same strategy. Table A.1 also shows the number of deadlocks that are removed or avoided. Note that several simulation runs just after each other can be removed in case of the first strategy and that several deadlocks can be avoided in one simulation run in case of the second strategy. This explains that the number of deadlocks can exceed 10 000 for both cases. The number of deadlocks that are avoided coincides with the number of dispatching interventions that are necessary to avoid these deadlocks with a first come first serve strategy. The upper part of Table A.1 can be regarded as a complement of Table 4.12.

Actually, the results in the *'Remove'* column are the same as the results in Table 4.12. The lower part complements Table 4.15.

Table A.1: Passenger robustness and deadlocks when using two strategies to handle deadlocks in the simulation. The upper part complements Table 4.12. The lower part complements Table 4.15.

	Passenger robustness ($\cdot 10^6$ min)		Impro (%)		# Deadlocks	
Timetable	Remove	Avoid	Remove	Avoid	Remove	Avoid
Infrabel	2.89	2.90	0.0	0.0	1 687	1 439
Dewilde	2.61	2.67	9.4	8.0	5 174	3 714
0% ds - 3h	3.33	3.40	-15.3	-17.1	47 361	17 490
25% ds - 3h	2.73	2.87	5.3	1.1	27 229	12 913
41% ds - 3h	2.56	2.59	11.4	10.7	6 796	7 024
50% ds - 3h	2.78	2.84	3.8	2.2	49 056	15 688
41% ds - 24h	2.63	2.65	8.8	8.8	4 696	3 658
iter 0	2.66	2.75	8.0	5.2	39 377	15 241
95% - 1	2.38	2.52	17.6	13.4	26 651	14 515
75% - 1	2.84	2.98	1.5	-2.4	17 635	8 383
75% - 2	2.63	2.69	8.9	7.3	22 140	13 765
50% - 1	2.57	2.63	10.9	9.3	17 797	9 975
50% - 2	2.69	2.79	6.7	4.1	40 698	14 132
50% - 3	3.06	2.95	-5.7	-1.7	31 955	13 786

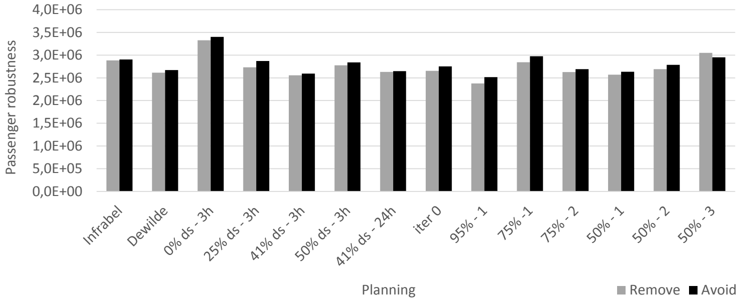


Figure A.1: Passenger robustness for two strategies to handle deadlocks in a simulation.

One can observe that the passenger robustness is generally higher (and thus worse) when the simulation deals with (avoid) the deadlocks instead of ignores (removes) the deadlocks. This is visualized in Figure A.1, that plots the passenger robustness obtained with both strategies for each timetable in Table A.1. The percentage difference in passenger robustness of the second strategy compared to the first strategy is also visualized in Figure A.2.

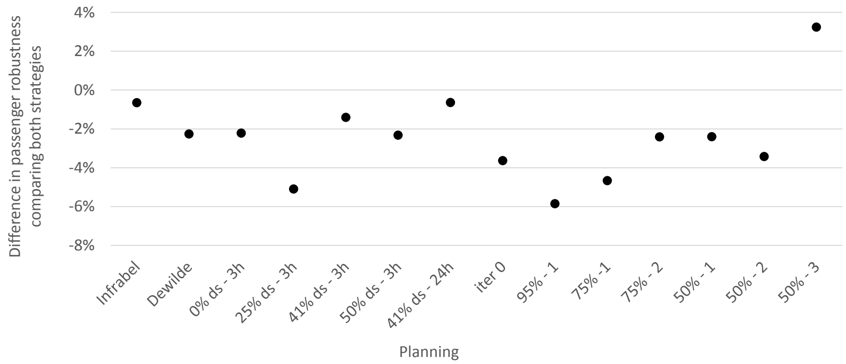


Figure A.2: Percentage difference in passenger robustness for avoiding deadlocks in simulation instead of ignoring simulation runs with a deadlock.

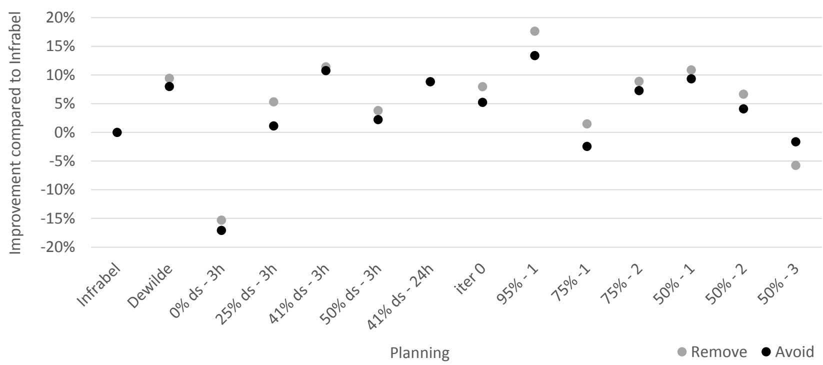


Figure A.3: Percentage improvement in passenger robustness for two strategies to handle deadlocks in simulation. The improvement is compared to the reference timetable of Infrabel, which is evaluated with the same strategy.

Furthermore, one can observe that the trends in percentage improvement compared to Infrabel is the same for both strategies. Figure A.3 plots this percentage improvement in passenger robustness compared to the reference timetable of Infrabel. The results of each strategy are compared to the evaluation of Infrabel's timetable with this same strategy. The two lines follow a similar pattern. However, the percentage improvement is a bit lower for the second strategy that avoids deadlocks.

Note that there are many more strategies to handle deadlocks in simulation. Rerouting, for example, is another strategy to avoid deadlocks, which is not considered here. We expect that the trends in improvement are independent of the dispatching strategy, but of course the final value of obtained passenger robustness will differ from strategy to strategy.

Table A.2 presents the amount of supplements, the average amount of non-used supplements and the average amount of knock-on delays when simulating the timetables in the first column while using the two different strategies to handle deadlocks. Again, the results in the 'Remove' column have been presented in Tables 4.12 and 4.15. We see that the amount of non-used supplements is slightly lower in case deadlocks are avoided and that the amount of knock-on delays is slightly higher. This can be explained by the fact that a deadlock is

Table A.2: Supplements and knock-on delays when using two strategies to handle deadlocks in the simulation. The upper part complements Table 4.12. The lower part complements Table 4.15.

Timetable	Supplement (min)	Non-used supplement (min) Remove	Non-used supplement (min) Avoid	Knock-on delay (min) Remove	Knock-on delay (min) Avoid
Infrabel	203.2	121.9	121.4	150.4	152.7
Dewilde	203.0	139.9	139.0	87.3	93.5
0% ds - 3h	0.0	0.0	0.0	179.0	190.9
25% ds - 3h	117.9	46.6	43.8	129.9	154.5
41% ds - 3h	205.7	105.1	102.9	117.3	126.2
50% ds - 3h	225.5	110.6	105.2	152.2	163.9
41% ds - 24h	205.7	104.9	103.4	122.8	125.8
iter 0	124.3	61.4	58.6	105.3	120.7
95% - 1	177.5	100.5	94.9	83.0	108.2
75% - 1	133.3	57.3	56.0	149.1	164.5
75% - 2	171.3	83.1	80.3	127.5	138.4
50% - 1	119.2	53.3	51.1	102.2	114.1
50% - 2	127.6	52.6	49.3	120.0	137.2
50% - 3	128.4	50.8	49.1	178.3	168.3

avoided by stopping one train before the shared infrastructure on which the deadlock would occur. This train has to wait there until the other train leaves the shared infrastructure, so it uses its supplement on that part of the network and the train gets a knock-on delay if its supplement is not sufficient.

Table A.3 presents the percentage of trains that leaves the network with a delay and the percentage of trains that recovers from its initial delay by using its supplements. These results are obtained by simulating the timetables in the first column and using the two different strategies to handle deadlocks. We see that the both performance indicators deteriorate if deadlocks are avoided instead of ignored. This can again be explained by the fact that avoiding a deadlock causes trains to wait such that trains cannot recover from initial delays and such that trains are (further) delayed.

Table A.3: Percentage trains that leave the network with a delay and the percentage of trains that recover from their initial delay when using two strategies to handle deadlocks in the simulation. The upper part complements Table 4.12. The lower part complements Table 4.15.

	% Delayed trains (%)		% Recovered trains (%)	
Timetable	Remove	Avoid	Remove	Avoid
Infrabel	60.2	61.4	9.7	9.1
Dewilde	49.2	50.9	12.8	11.9
0% ds - 3h	80.8	81.0	0.0	0.0
25% ds - 3h	59.7	63.5	10.6	9.6
41% ds - 3h	43.0	44.6	18.3	17.8
50% ds - 3h	46.0	49.8	18.8	17.4
41% ds - 24h	41.8	42.8	18.9	18.5
iter 0	55.7	59.0	10.1	8.8
95% - 1	48.2	54.9	15.6	13.8
75% - 1	60.9	64.2	10.6	10.1
75% - 2	53.9	56.1	14.0	13.3
50% - 1	54.5	57.6	11.6	10. 7
50% - 2	55.1	58.4	12.7	11.8
50% - 3	57.0	59.9	12.7	11.8

Figure A.4 shows that there is no clear correlation between the number of simulation runs that are thrown away in case that deadlocks are ignored and the difference in percentage improvement in passenger robustness compared to the reference timetable of Infrabel. The vertical axis presents the number of deadlocks and the horizontal axis presents the difference in percentage improvement in passenger robustness compared to the reference timetable of Infrabel for both strategies to handle delays. In order to understand the value on the horizontal axis, it is now explained for the timetable of Dewilde. This value is obtained as $9.4\% - 8.0\% = 1.4\%$, where 9.4% is the percentage improvement in passenger robustness of Dewilde compared to Infrabel in case deadlocks are ignored and 8.0% is the percentage improvement of Dewilde compared to Infrabel in case deadlocks are avoided. We omitted the observation for timetable '50% - 3' in Figure A.4, since this is the only timetable where the passenger robustness is better in case deadlocks are avoided instead of ignored, which gives a negative difference in percentage improved passenger robustness compared to the reference timetable of Infrabel.

Figure A.4 shows that the number of simulation runs that are thrown away in case deadlocks are ignored, cannot be used to forecast the relative performance compared to the timetable of Infrabel in case deadlocks are avoided.

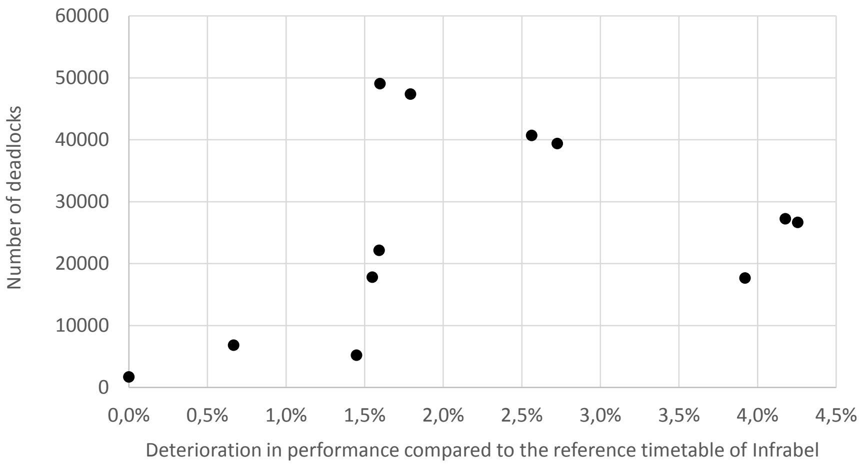


Figure A.4: The vertical axis presents the number of deadlocks and the horizontal axis presents the difference in percentage improvement in passenger robustness compared to the reference timetable of Infrabel for both strategies to handle delays.

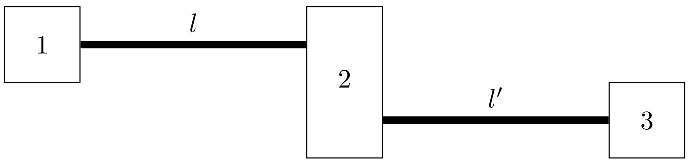
A.1.3 Conclusion and future work

There exist many ways to handle deadlocks in simulation. In general, ignoring simulation runs in which a deadlock occurs gives an optimistic value for the passenger robustness compared to a strategy where deadlocks are avoided with a simple first come first serve strategy. The trends in improvement and the conclusions of our research, however, remain the same whether simulation runs with deadlocks are ignored or deadlocks are avoided with a first come first serve strategy, but the absolute improvement is a bit smaller for the latter strategy. Avoiding deadlocks in the simulation is probably a better representation of practice. However, it should be noted that in this case the presented results not only evaluate the performance of a given timetable, but also implicitly the dispatching strategy implemented in the simulation to avoid the deadlocks. Probably better dispatching strategies are implemented in practice. Furthermore, there is no clear correlation between the number of simulation runs that are thrown away in case that deadlocks are ignored and the difference in percentage improvement in passenger robustness compared to the reference timetable of the Belgian railway infrastructure manager Infrabel.

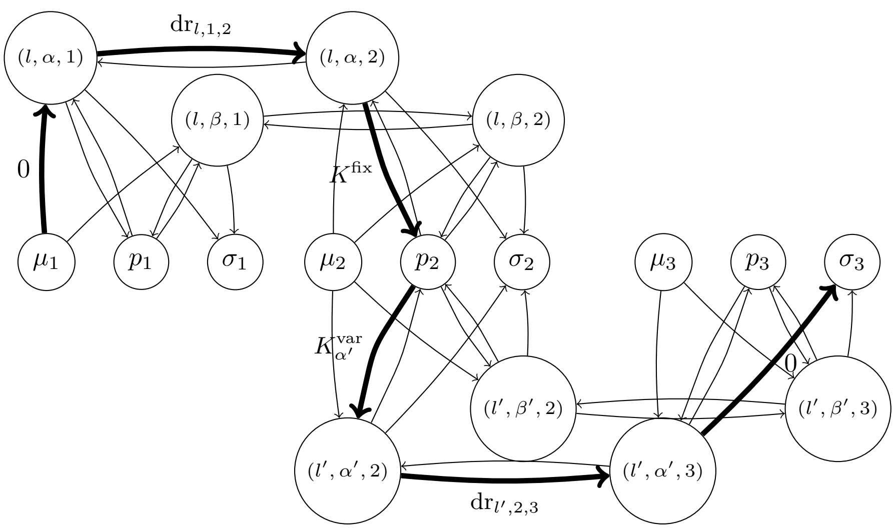
Tracking which trains cause deadlocks and at which locations deadlocks are caused could give information on how a timetable can be further improved. It is interesting to investigate these data together with the delays upon arrival in the corresponding simulation runs. Another idea is to look at how a knock-on delay caused by a deadlock spreads itself over the network and the trains.

A.2 Passenger graph

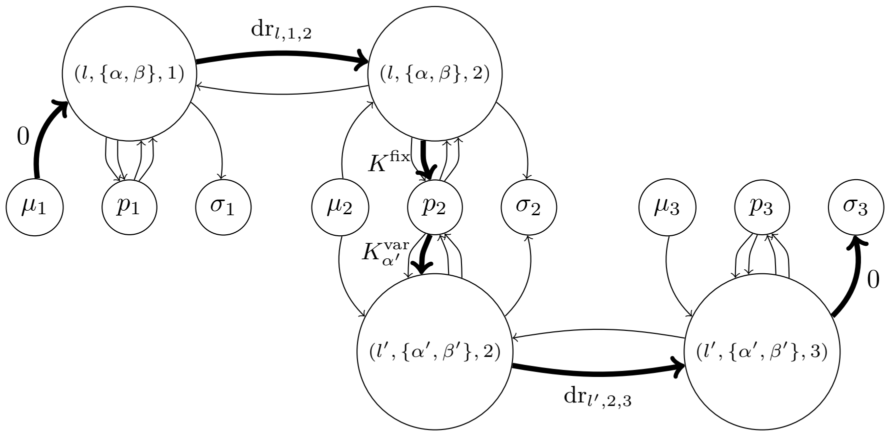
This appendix shows a more complex example of a passenger graph in which multiple frequencies per line are visualized. Like in Figure 5.4, costs are labeled on the edges for a passenger traveling from station 1 to station 3, transferring lines at station 2, with used edges in bold; considering that, in Figure A.5c, line l' operates at frequency α'.



(a) Line structure (two lines, three stations).



(b) Full graph structure.



(c) Reduced graph structure.

Figure A.5: The upper figure shows a simple network with three stations 1, 2 and 3, and two lines l and l'. Line l visits stations 1 and 2, and line l' visits stations 2 and 3. Line l and l' are each operating at two different frequencies, $\alpha, \beta \in \mathbb{Z}^+$ and $\alpha', \beta' \in \mathbb{Z}^+$ respectively. We show the graph structure before and after an aggregation of the different frequency components for a line. The corresponding (line, frequency, station) vertices are collapsed, leading to multi-edges where the frequency-dependent costs must be considered.

A.3 List of symbols

A.3.1 Symbols of Chapter 3

Sets	
R_t	Set of all possible routes for train t
$R^c_{(t,r),t'}$	Set of all routes for train t' conflicting with (t,r)
T	Set of trains
Specific elements of a set	
r, r' or r''	Route
s_i or s_j	Railway network section
t, t' or A, B	Train
(t,r)	Train-route combination of train t and route r
Characteristics and properties of elements	
$B_{t,t'}$	Planned minimum buffer time between train t and t'
$B_{t,r,t',r'}$	Planned minimum buffer time between (t,r) and (t',r')
$C_{t,t'}$	Cost associated with $B_{t,t'}$
$C_{t,r,t',r'}$	Cost associated with $B_{t,r,t',r'}$
$C^P_{t,r,t',r'}$	Passenger weighted cost associated with $B_{t,r,t',r'}$
$C^D_{t,r,t',r'}$	Delay weighted cost associated with $B_{t,r,t',r'}$
D_t	Mean recurring delay of train t
$D_{t,t'}$	Weight w.r.t. the mean recurring delay of train t and t'
$P_{t,t'}$	Passenger weight w.r.t. train t and t'
$P_{t,r,t',r'}$	Passenger weight w.r.t. (t,r) and (t',r')
run-sup_{s_i}	Running time supplement in section s_i
Decision variables	
$x_{t,r}$	Binary train-route combination variable
$z_{t,r}$	Spreading cost associated with train t and route r

A.3.2 Symbols of Chapter 4

Sets	
$\mathcal{P}$	Set of platforms in a platform area
R	Set of routes
R_t	Set of all possible routes of train t
$\mathcal{R}_{\vec{t},\mathrm{dep},w}$	Set of partial routes with direction $\vec{t}$ departing from node w
$\mathcal{R}_{\vec{t},\mathrm{arr},w}$	Set of partial routes with direction $\vec{t}$ arriving in node w
T	Set of trains
T_w	Set of trains for which r^t contains node w
$\mathcal{V}_p$	Set of transfer facilitating platforms w.r.t. platform p
W	Set of nodes
Specific elements of a set	
B	Buffer time
e^{ℓ_t}	Last node of the line operated by train t
k	Indicator related to the number of node usages
p, p'	Platform
r, r', r_i	(Partial) route
r^t	Route that is assigned to train t
$s_r\ (s_{r^t})$	First node on route r (r^t)
s^{ℓ_t}	First node of the line operated by train t
t, t', t'', t_i	Train
w, w_i	Node
Characteristics and properties of elements	
c_k	Linearization coefficients in the routing model
$\mathrm{group}(B)$	Buffer time group of buffer time B w.r.t. possibly affected passengers
$l_{r,w}$	Parameter indicating whether node w lies on route r
$q_{t,r,w}$	Time that train t on route r blocks node w
$\vec{t}$	Direction of train t
Parameters	
C	Lowest common multiple of $\{1, \cdots J\}$
$d^{\mathrm{res}}_{t,r^t,w}$	Duration w.r.t. train t, route r and the reservation of node w
$d^{\mathrm{rel}}_{t,r^t,w}$	Duration w.r.t. train t, route r and the release of node w
$\mathcal{H}$	Objective weight in the routing model
$\tilde{\mathcal{H}}$	Objective weight in the timetabling model
L	Lower bound for the transfer time
$\lambda_{t_i,t_j,w}$	Passenger weight of trains t_i and t_j in node w
J	Number of buffer time groups w.r.t.

	possibly affected passengers
K	Upper bound on the maximum node usage
ξ	Maximum number of periods necessary in an auxiliary timetable
P	Period length of the cyclic timetable
$\lvert R\rvert$	Number of routes in set R
$\lvert T\rvert$	Number of trains in set T
U	Upper bound for the transfer time
$\lvert W\rvert$	Number of nodes in set W
Decision variables	
$b_{k,w}$	Linearization variable in the routing model w.r.t. node w
$\mathrm{bin}_{p,t}$	Platform decision variable for train t
$\mathrm{buf}_{t_i,t_j,w}$	Buffer time between trains t_i and t_j in node w
g_w	Number of times that node w is used
$\mathrm{minbuf}_{t_i,t_j}$	Minimum buffer time between trains t_i and t_j
$\mathrm{res}_{t,w}$	Reservation time of node w by train t
$\mathrm{rel}_{t,w}$	Release time of node w by train t
$\mathrm{res}_{t,w}^{\mathrm{aux}}$	Auxiliary reservation time of node w by train t
$\mathrm{rel}_{t,w}^{\mathrm{aux}}$	Auxiliary release time of node w by train t
$\mathrm{res}_{t,w}^{\mathrm{int}}$	Modulo variable related to $\mathrm{res}_{t,w}$
$\mathrm{rel}_{t,w}^{\mathrm{int}}$	Modulo variable related to $\mathrm{rel}_{t,w}$
$\mathrm{resrel}_{t_i,t_j,w}^{\mathrm{int}}$	Modulo variable related to $\mathrm{res}_{t_j,w}$ and $\mathrm{rel}_{t_i,w}$
$\mathrm{relres}_{t_i,t_j,w}^{\mathrm{int}}$	Modulo variable related to $\mathrm{rel}_{t_j,w}$ and $\mathrm{res}_{t_i,w}$
$\mathrm{relrel}_{t_i,t_j,w}^{\mathrm{int}}$	Modulo variable related to $\mathrm{rel}_{t_j,w}$ and $\mathrm{rel}_{t_i,w}$
$x_{t,r}$	Binary route decision variable for train t and route r
z	Overall maximum node usage
$\tilde{z}$	Overall minimum buffer time

A.3.3 Symbols of Chapter 5

Sets	
$\mathcal{A}$	Activity set
$\mathcal{A}^{\text{run}}$	Set of running activities
$\mathcal{A}^{\text{station}}$	Set of station activities
$\mathcal{A}^{\text{buffer}}$	Set of buffer activities
$\mathcal{A}^{\text{turn}}$	Set of turn activities
$\mathcal{A}^{\text{line spread}}$	Set of line spread activities
$\mathcal{A}^{\text{turn-con}}$	Set of turn connection activities
$\mathcal{E}$	Event set
$\mathcal{E}^{\text{res}}$	Set of reservation events
$\mathcal{E}^{\text{res},p}$	Set of platform reservation events
$\mathcal{E}^{\text{rel}}$	Set of release events
$\mathcal{E}^{\text{rel},p}$	Set of platform release events
$(\mathcal{E}, \mathcal{A})$	Event-activity network
E	Edge set of the passenger graph
$\mathcal{F}_l$	Potential frequencies of line l
$\mathcal{L}$	Line pool
$\mathcal{L}_r$	Subset of lines in $\mathcal{L}$ that make use of resource r
$\mathcal{R}$	Infrastructure resources
$\mathcal{S}$	Station (area) set
$\mathcal{S}_l$	Set of stations on line l
T	Set of trains
$T_{\text{line spread}}$	Set of train couples considered for line spreading
T_{turn}	Set of train couples considered for turning
V	Vertex set of the passenger graph
$\mathcal{X}$	Line plan solution
Specific elements of a set	
e	Edge
l or l'	Line
r	(Infrastructure) resource
s or s'	Station
t or t'	Train
v	Vertex
X, Y	Terminal stations
Characteristics and properties of elements	
a_v^s	Flow of passengers from station s in vertex v
C_f	Capacity of a line operating at frequency f
$c_{l,f}$	Operator cost of line l at frequency f

$d_{s,s'}$	Demand between stations s and s'
$\mathrm{dr}_{l,s,s'}$	Driving time of line l from station s to station s'
f_l	The frequency of line l
f_e	Frequency of edge e
l_e	Line of edge e
ℓ_t	Line operated by train t
L_a	Lower bound for activity a
$\mu(_s)$	Source vertex (of station s) in the passenger graph
ntt_s	Necessary turn time in station s
$\mathrm{occ}_{l,s}$	Occupation time of line l in station s
$p(_s)$	Platform vertex (of station s) in the passenger graph
$\rho_{\tilde{s}_t,t}$	Platform of train t in terminal station $\tilde{s}_t$
rmin_r	Min resource usage of resource r
rmax_r	Max resource usage of resource r
$\mathrm{run}_{l,s,s'}$	Running time of line l between stations s and s'
$\sigma(_s)$	Sink vertex (of station s) in the passenger graph
$s_{l,i}$	The i-th station of line l
$\tilde{s}_t$	Terminal station of train t, i.e. $s_{\ell_t,\lvert\mathcal{S}_{\ell_t}\rvert}$
τ_e	Passenger cost for edge e of the passenger graph
$\mathcal{T}_l$	Travel time on line l from begin to end station
$t^i_{l,r}$	The i-th train of line l considered on resource r
U_a	Upper bound for activity a
φ_p	Usage frequency of platform p
Parameters	
ϵ	Time discretization
B	Minimum necessary buffer time
λ	Weight for the operator cost in the line planning objective
M	Minimum planned buffer time between two lines
P	Period length of the cyclic timetable
Decision variables	
k_a	Variable to induce a positive activity time for activity a
π_i	Event time of event i
$x_{l,f}$	Binary line decision variable of line l at frequency f
y^e_s	Flow decision variable of station s along edge e
z	Minimum overall buffer time

Bibliography

UIC. In *International Union of Railways*. Second edition, 2013. Code 406: Capacity.

Ravindra K. Ahuja, Rolf H. Möhring, and Christos Zaroliagis, editors. *Robust and Online Large-Scale Optimization*, volume 5868. Springer-Verlag, Heidelberg, 2009.

Emma V. Andersson. Assessment of robustness in railway traffic timetables. Master's thesis, Linköping University, Sweden, 2014.

Emma V. Andersson, Anders Peterson, and Johanna Törnquist Krasemann. Quantifying railway timetable robustness in critical points. *Journal of Rail Transport Planning and Management*, 3(3):95–110, 2013.

Emma V. Andersson, Anders Peterson, and Johanna Törnquist Krasemann. Improved railway timetable robustness for reduced traffic delays - a MILP approach. In *Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015)*, 2015.

Lukas Bach, Twan Dollevoet, and Dennis Huisman. Integrating timetabling and crew scheduling at a freight railway operator. *Transportation Science*, 50 (3):878–891, 2016.

Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 1:35–53, 2004.

Nikola Bešinović, Rob M.P. Goverde, Egidio Quaglietta, and Roberto Roberti. An integrated micro-macro approach to robust railway timetabling. *Transportation Research Part B: Methodological*, 87:14–32, 2016a.

Nikola Bešinović, Rob M.P. Goverde, and Egidio Quaglietta. Microscopic models and network transformations for automated railway traffic planning. *Computer-Aided Civil and Infrastructure Engineering*, 32(2):89–106, 2017. doi: 10.1111/mice.12207.

Ralf Borndörfer, Martin Grötschel, and Marc E. Pfetsch. A column-generation approach to line planning in public transport. *Transportation Science*, 41(1): 123–132, 2007.

Sofie Burggraeve and Pieter Vansteenwegen. Optimization of supplements and buffer times in passenger robust timetabling. In *Proceedings of the 7th International Seminar on Railway Operations Modelling and Analysis (RailLille2017), Lille, France. 8th best paper of the RailLille2017 conference*, 2017a.

Sofie Burggraeve and Pieter Vansteenwegen. Optimization of supplements and buffer times in passenger robust timetabling. *Journal of Rail Transport Planning & Management*, In Press, 2017b. URL `https://doi.org/10.1016/j.jrtpm.2017.08.004`.

Sofie Burggraeve and Pieter Vansteenwegen. Robust routing and timetabling in complex railway stations. *Transportation Research Part B: Methodology*, 101: 228–244, 2017c.

Sofie Burggraeve, Thijs Dewilde, Tim Ameye, Peter Sels, and Pieter Vansteenwegen. Definitie van robuustheid voor treinsystemen (Defining robustness for railway systems). *Tijdschrift Vervoerswetenschap*, 51(1):49–66, 2015a.

Sofie Burggraeve, Thijs Dewilde, Peter Sels, and Pieter Vansteenwegen. Improving passenger robustness by taking passenger numbers and recurring delays explicitly into account on the tactical level. In *Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015), Tokyo, Japan*, 2015b.

Sofie Burggraeve, Sofie Van Thielen, and Pieter Vansteenwegen. Advanced methods for improving passenger robustness in railway systems. In *Proceedings of the Conference on Advanced Systems in Public Transport (CASPT2015), Rotterdam, The Netherlands*, 2015c.

Sofie Burggraeve, Simon H. Bull, Richard M. Lusby, and Pieter Vansteenwegen. Train turn restrictions and line plan performance. In *Proceedings of the Ninth*

Triennial Symposium on Transportation Analysis (TRISTAN IX), Oranjestad, Aruba, 2016.

Sofie Burggraeve, Simon H. Bull, Pieter Vansteenwegen, and Richard M. Lusby. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160, 2017.

Michael R. Bussieck, Peter Kreuzer, and Uwe T. Zimmermann. Optimal lines for railway systems. *European Journal of Operational Research*, 96(1):54–63, 1997.

Valentina Cacchiani and Paolo Toth. Nominal and robust train timetabling problems. *European Journal of Operational Research*, 219:727–737, 2012.

Valentina Cacchiani, Fabio Furini, and Martin P. Kidd. Approaches to a real-world train timetabling problem in a railway node. *Omega*, 58:97–110, 2016.

Gabrio Caimi, Fabian Chudak, Martin Fuchsberger, Marco Laumanns, and Rico Zenklusen. A new rescource-constrained multicommodity flow model for conflict-free train routing and scheduling. *Transportation Science*, 45(2): 212–227, 2011.

Gabrio Caimi, Martin Fuchsberger, Marco Laumanns, and Marco Luethi. A model predictive control approach for discrete-time rescheduling in complex central railway station areas. *Computers & Operations Research*, 39:2578–2593, 2012.

Gabrio C. Caimi. *Algorithmic decision support for train scheduling in a large and highly utilised railway network*. PhD thesis, Eidgenössische Technische Hochschule ETH Zürich, 2009. Nr. 18581.

Gabrio C. Caimi, Martin Fuchsberger, Dan Burkolter, Thomas Hermann, Raimond Wüst, and Samuel Roos. Conflict-free train scheduling in a compensation zone exploiting the speed profile. In *Proceedings of the 3rd International Seminar on Railway Operations Modelling and Analysis (RailZurich2009), Zurich, Switzerland*, 2009a.

Gabrio C. Caimi, Thomas Hermann, Dan Burkolter, Fabian Chudak, and Marco Laumanns. Design of a railway scheduling model for dense services. *Networks and Spatial Economics*, 9(1):25–46, 2009b.

Alberto Caprara, Matteo Fischetti, and Paolo Toth. Modelling and solving the train timetabling problem. *Operations Research*, 50:851–861, 2002.

Malachy Carey and David Lockwood. A model, algorithms and strategy for train pathing. *Journal of the Operational Research Society*, 46:988–1005, 1995.

Serafino Cicerone, Gianlorenzo D'Angelo, Gabriele Di Stefano, Daniele Frigioni, and Alfredo Navarra. Recoverable robust timetabling for single delay: Complexity and polynomial algorithms for special cases. *Journal of Combinatorial Optimization*, 18:2–29, 2009.

Stefano De Fabris, Giovanni Longo, Giorgio Medeossi, and Raffaele Pesenti. Application and validation of a timetabling algorithm to a large Italian network. In *Proceedings of the 5th International Seminar on Railway Operations Modelling and Analysis (RailCopenhagen2013), Kgs. Lyngby, Denmark*, 2013.

Alicia De-Los-Santos, Gilbert Laporte, Juan A. Mesa, and Federico Perea. Evaluating passenger robustness in a rail transit network. *Transportation Research Part C: Emerging Technologies*, 20(1):34–46, 2011.

Thijs Dewilde. *Improving the robustness of a railway system in large and complex station areas.* PhD thesis, KU Leuven, Belgium, 2014b.

Thijs Dewilde, Peter Sels, Dirk Cattrysse, and Pieter Vansteenwegen. Defining robustness of a railway timetable. In *Proceedings of the 4th International Seminar on Railway Operations Modelling and Analysis (RailRome2011), Rome, Italy*, 2011.

Thijs Dewilde, Peter Sels, Dirk Cattrysse, and Pieter Vansteenwegen. Robust railway station planning: An interaction between routing, timetabling and platforming. *Journal of Rail Transport Planning & Management*, 3(3):68–77, 2013.

Thijs Dewilde, Peter Sels, Dirk Cattrysse, and Pieter Vansteenwegen. Improving the robustness in railway station areas. *European Journal of Operational Research*, 235(1):276–286, 2014a.

Ophelia Engelhardt-Funke and Michael Kolonko. Analysing stability and investments in railway networks using advanced evolutionary algorithms. *International Transactions in Operational Research*, 11:381–394, 2004.

Matteo Fischetti, Domenico Salvagnin, and Arrigo Zanette. Fast approaches to improve the robustness of a railway timetable. *Transportation Science*, 43(3): 321–335, 2009.

Philine Gattermann, Peter Großmann, Karl Nachtigall, and Anita Schöbel. A SAT-formulation for integrating line planning, timetabling and passenger routing. In *Proceedings of the International Conference on Operations Research of the German Operations Research Society (GOR), Hamburg, Germany*, 2016.

Marc Goerigk, Michael Schachtebeck, and Anita Schöbel. Evaluating line concepts using travel times and robustness. *Public Transport*, 5(3):267–284, 2013.

Eliyahu M. Goldratt and Jeff Cox. The goal: A process of ongoing improvement. 1986.

Jan-Willem Goossens, Stan van Hoesel, and Leo Kroon. On solving multi-type railway line planning problems. *European Journal of Operational Research*, 168(2):403–424, 2006.

Rob Goverde. Optimal scheduling of connections in railway systems. In *Proceedings of the 8th WCTR, Antwerp, Belgium*, 1998.

Rob M.P. Goverde, Nikola Bešinović, Anne Binder, Valentina Cacchiani, Egidio Quaglietta, Roberto Roberti, and Paolo Toth. A three-level framework for performance-based railway timetabling. *Transportation Research Part C: Emerging Technologies*, 67:62–83, 2016.

Peter Großmann. Polynomial reduction from PESP to SAT. Technical report, TU Dresden, 2011.

Tore Grünert, Stefan Irnich, Hans-Jürgen Zimmermann, Markus Schneider, and Burkhard Wulfhorst. Finding all k-cliques in k-partite graphs, an application in textile engineering. *Computers & operations research*, 29:13–31, 2002.

Ingo A. Hansen and Jörn Pachl, editors. *Railway timetable & traffic: Analysis, modelling and simulation.* Eurailpress, Hamburg, 2008.

Dennis Huisman, Leo G. Kroon, Ramon M. Lentink, and Michiel J.C.M. Vromans. Operations research in passenger railway transportation. *Statistica Neerlandica*, 59:467–497, 2005.

Lars W. Jensen and Alex Landex. Measuring robustness of timetables at stations using a probability distribution. In *Proceedings of the 5th International Seminar on Railway Operations Modelling and Analysis, RailCopenhagen, Kgs. Lyngby, Denmark*, 2013.

Predrag Jovanović, Pavle Kecman, Nebojša Bojović, and Dragomir Mandić. Optimal allocation of buffer times to increase train schedule robustness. *European Journal of Operational Research*, 256(1):44–54, 2017.

Mor Kaspi and Tal Raviv. Service-oriented line planning and timetabling for passenger trains. *Transportation Science*, 47(3):295–311, 2013.

Leonard Kaufman. An algorithm for the quadratic assignment problem using Bender's decomposition. *European Journal of Operational Research*, 2(3): 207–211, 1978.

Muhammad B. Khan and Xuesong Zhou. Stochastic optimization model and solution algorithm for robust double-track train-timetabling problem. *IEEE Transactions on Intelligent Transportation Systems*, 11(1):81–89, 2010.

Leo Kroon, Gábor Maróti, Mathijn Retel Helmrich, Michiel Vromans, and Rommert Dekker. Stochastic improvement of cyclic railway timetables. *Transportation Research Part B: Methodological*, 42:553–570, 2008a.

Leo G. Kroon, Rommert Dekker, and Michiel J.C.M Vromans. Cyclic railway timetabling: a stochastic optimization approach. In *Lecture Notes in Computer Science*, volume 4359, pages 41–66. 2007b.

Leo G. Kroon, Dennis Huisman, and Gábor Maróti. Optimisation models for railway timetabling. In Ingo A. Hansen and Jörn Pachl, editors, *Railway timetable & traffic: Analysis, modelling and simulation*, pages 135–154. Eurailpress, Hamburg, 2008b.

Leo G. Kroon, Dennis Huisman, Erwin Abbink, Pieter-Jan Fioole, Matteo Fischetti, Gábor Maróti, Alexander Schrijver, Adri Steenbeek, and Roelof Ybema. The new dutch timetable: The OR revolution. *Interfaces*, 39(1): 6–17, 2009.

Leonardo Lamorgese, Carlo Mannino, and Erik Natvig. An exact micro-macro approach to cyclic and non-cyclic train timetabling. *Omega*, 2016. URL http://dx.doi.org/10.1016/j.omega.2016.11.004.

Gilbert Laporte, Juan A. Mesa, and Federico Perea. A game theoretic framework for the robust railway transit network design problem. *Transportation Research Part B: Methodological*, 44:447–459, 2010.

Tomas Lidén and Martin Joborn. An optimization model for integrated planning of railway traffic and network maintenance. *Transportation Research Part C: Emerging technologies*, 74:327–347, 2017.

Christian Liebchen. Symmetry for periodic railway timetables. *Electronic Notes in Theoretical Computer Science*, 92(1):1–17, 2003.

Christian Liebchen. *Periodic Timetable Optimization in Public Transport*. PhD thesis, Technische Universität Berlin, 2006.

Christian Liebchen and Rolf H. Möhring. The modeling power of the periodic event scheduling problem: railway timetables - and beyond. In Frank Geraets, Leo Kroon, Anita Schöbel, Dorothea Wagner, and Christos D. Zaroliagis, editors, *Algorithmic methods for railway optimization*, pages 3–40. Springer, 2007.

Christian Liebchen, Michael Schachtebeck, Anita Schöbel, Sebastian Stiller, and André Prigge. Computing delay resistant railway timetables. *Computers & Operations Research*, 37:857–868, 2010.

Richard M. Lusby, Jesper Larsen, Matthias Ehrgott, and David Ryan. Railway track allocation: Models and methods. *OR Spectrum*, 33(4):843–883, 2011.

Richard M. Lusby, Jesper Larsen, and Simon Bull. A survey on robustness in railway planning. *European Journal of Operational Research*, In Press, 2017. URL `https://doi.org/10.1016/j.ejor.2017.07.044`.

Ludolf E. Meester and Sander Muns. Stochastic delay propagation in railway networks and phase-type distributions. *Transportation Research Part B: Methodological*, 41(2):218–230, 2007.

Mathias Michaelis and Anita Schöbel. Integrating line planning, timetabling, and vehicle scheduling: a customer-oriented heuristic. *Public Transport*, 1(3): 211–232, 2009. ISSN 1613-7159.

Elise D. Miller-Hooks and Hani S. Mahmassani. Least expected time paths in stochastic, time-varying transportation networks. *Transportation Science*, 34 (2):198–215, 2000.

Karl Nachtigall. Periodic network optimization with different arc frequencies. *Discrete Applied Mathematics*, 69:1–17, 1996.

OPEN TRACK. URL `http://www.opentrack.ch`. Accessed: August 18, 2017.

Jens Parbo, Otto A. Nielsen, and Carlo G. Prato. Passenger perspectives in railway timetabling: A literature review. *Transport Reviews*, pages 1–27, 2016. URL `http://dx.doi.org/10.1080/01441647.2015.1113574`.

Leon W.P. Peeters. *Cyclic Railway Timetable Optimization*. PhD thesis, Erasmus Research Institute of Management (ERIM), 2003.

Anders Peterson. Towards a robust traffic timetable for the Swedisch southern mainline. *WIT Transactions on the built environment*, 127:473–484, 2012.

Helena Pieters. Een heuristiek voor het opstellen van een dienstregeling voor het S-tog netwerk in kopenhagen. Master's thesis, KU Leuven, Belgium, 2017.

Natalia J. Rezanova. Line planning optimization at DSB. In *Proceedings of the Conference on Advanced Systems in Public Transport*, Rotterdam, The Netherlands, 2015.

Ricardo Rail. Realistic railway network part. URL `https://magnet.me/a/company/ricardo-rail/about`. Accessed: May 23, 2017.

Raphaela Rudolph. Allowances and margins in railway scheduling. In *Proceedings of WCRR*, 2003.

Miguel A. Salido, Federico Barber, and Laura Ingolotti. Robustness for a single railway line: Analytical and simulation methods. *Expert Systems with Applications*, 39(18):13305 – 13327, 2012. ISSN 0957-4174.

Thomas Schlechte, Ralf Borndörfer, Berkan Erol, Thomas Graffagnino, and Elmar Swarat. Micro-macro transformation of railway networks. *Journal of Rail Transport Planning & Management*, 1(1):38–48, 2011.

Marie Schmidt and Anita Schöbel. Timetabling with passenger routing. *OR Spectrum*, 37(1):75–97, 2015.

Anita Schöbel. Line planning in public transportation: models and methods. *OR Spectrum*, 34(3):491–510, 2012.

Anita Schöbel. Integration of line planning and timetabling. In *Proceedings of the Conference on Advanced Systems in Public Transport*, Rotterdam, The Netherlands, 2015.

Anita Schöbel and Susanne Scholl. Line planning with minimal traveling time. In *ATMOS 2005-5th Workshop on Algorithmic Methods and Models for Optimization of Railways*. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, 2006.

Alexander Schrijver and Adri Steenbeek. Spoorwegdienstregelingontwikkeling (Timetable construction). Technical report, CWI Center for Mathematics, 1993.

Peter Sels. *Large-Scale, Passenger Oriented, Cyclic Railway Timetabling and Station Platforming and Routing*. PhD thesis, 2016b.

Peter Sels, Thijs Dewilde, Dirk Cattrysse, and Pieter Vansteenwegen. A passenger knock-on delay model for timetable optimisation. In T. Albrecht, B. Jaekel, and M. Lehnert, editors, *Proceedings of the 3rd International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 377–387, 2013.

Peter Sels, Pieter Vansteenwegen, Thijs Dewilde, Dirk Cattrysse, B. Waquet, and A. Joubert. The train platforming problem: The infrastructure management company perspective. *Transportation Research Part B: Methodological*, 61: 55–72, 2014.

Peter Sels, Dirk Cattrysse, and Pieter Vansteenwegen. Practical macroscopic evaluation and comparison of railway timetables. In *Transportation research procedia*, volume 10, pages 625–633, Delft, The Netherlands, 2015.

Peter Sels, Thijs Dewilde, Dirk Cattrysse, and Pieter Vansteenwegen. Reducing the passenger travel time in practice by the automated construction of a robust railway timetable. *Transportation Research Part B: Methodological*, 84:124–156, 2016a.

Paolo Serafini and Walter Ukovich. A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581, 1989.

Michael Siebert and Marc Goerigk. An experimental comparison of periodic timetabling models. *Computers & Operations Research*, 40(1):2251–2259, 2013.

Yahua Sun, Chengxuan Cao, and Chao Wu. Multi-objective optimization of train routing problem combined with train scheduling on a high-speed railway network. *Transportation Research Part C: Emerging Technologies*, 44:1–20, 2014.

Yoko Takeuchi, Norio Tomii, and Chikara Hirai. Evaluation method of robustness for train schedules. Technical report, 2007.

treinreiziger.nl. URL `http://www.treinreiziger.nl/superdagretour-goedkoop-treinkaartje-naar-belgie/`. Accessed: September 11, 2017.

Pieter Vansteenwegen. *Planning in tourism and public transportation.* PhD thesis, KU Leuven, 2008.

Pieter Vansteenwegen and Dirk Van Oudheusden. Decreasing the passenger waiting time for an intercity rail network. *Transportation Research Part B: Methodological*, 41:478–492, 2006.

Pieter Vansteenwegen, Thijs Dewilde, Sofie Burggraeve, and Dirk Cattrysse. An iterative approach for reducing the impact of infrastructure maintenance on the performance of railway systems. *European Journal of Operational Research*, 252(1):39–53, 2016.

Jianxin Yuan and Ingo A. Hansen. Optimizing capacity utilization of stations by estimating knock-on train delays. *Transportation Research Part B: Methodological*, 41:202–217, 2007.

Peter J. Zwaneveld, Leo G. Kroon, H. Edwin Romeijn, Marc Salomon, Stéphane Dauzère-Pérès, Stan P.M. van Hoesel, and Harrie W. Ambergen. Routing trains throug railway stations: Model formulation and algorithms. *Transportation Science*, 30:181–194, 1996.

Peter J. Zwaneveld, Leo G. Kroon, and Stan P.M. van Hoesel. Routing trains through a railway station based on a node packing model. *European Journal of Operational Research*, 128:14–33, 2001.

Curriculum vitae

www.linkedin.com/in/sofie-burggraeve-6aa37186

Education

PhD in Eng. Sc. 10/2013-09/2017	KU Leuven Mobility Research Centre PhD Topic: Passenger robust timetables for dense railway networks Supervisor: Pieter Vansteenwegen
Research stay abroad 09/2015-12/2015	Research stay at DTU Management Engineering, Copenhagen, Denmark
Specific Teacher Training Magna cum laude 2011-2014	KU Leuven Option: Mathematics
M.Sc. in Mathematics Summa cum laude 2011-2013	KU Leuven Option: Business Master Thesis: Tropical geometry, Linear systems on metric graphs Supervisor: Filip Cools

B.Sc. in Mathematics Cum laude 2008-2011	KU Leuven

Experience

Assistant in Mathematics 2011-2013	Student job at KU Leuven
Research Internship 2012	VIVES, KU Leuven Topic: Solutions to the lock congestion problem Supervisor: Jo Reynaerts
Tutor in Mathematics 2011	Rebus Studiebegeleiding
Praesidium member 2009-2011	Organization: Wina, the student union for mathematics, physics and computer science students at KU Leuven Function: financial manager and first year student support
Youth organization 1998-2010	Member and youth leader of scouts Diependaal

Languages

Dutch: native
English: very good
French, German: good

Extra information

International summer school July 10-18, 2011	Number Theory Gulbenkian Foundation, Lisbon

Publications

Articles in internationally reviewed academic journals

Sofie Burggraeve and Pieter Vansteenwegen. Robust routing and timetabling in complex railway stations. *Transportation Research Part B: Methodology*, 101: 228–244, 2017c.

Sofie Burggraeve and Pieter Vansteenwegen. Optimization of supplements and buffer times in passenger robust timetabling. *Journal of Rail Transport Planning & Management*, In Press, 2017b. URL https://doi.org/10.1016/j.jrtpm.2017.08.004.

Sofie Burggraeve, Simon H. Bull, Pieter Vansteenwegen, and Richard M. Lusby. Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160, 2017.

Pieter Vansteenwegen, Thijs Dewilde, Sofie Burggraeve, and Dirk Cattrysse. An iterative approach for reducing the impact of infrastructure maintenance on the performance of railway systems. *European Journal of Operational Research*, 252(1):39–53, 2016.

Articles in other academic journals

Sofie Burggraeve, Thijs Dewilde, Tim Ameye, Peter Sels, and Pieter Vansteenwegen. Definitie van robuustheid voor treinsystemen (Defining robustness for railway systems). *Tijdschrift Vervoerswetenschap*, 51(1):49–66, 2015a.

Papers at international scientific conferences and symposia, published in full proceedings

Sofie Burggraeve and Pieter Vansteenwegen. Optimization of supplements and buffer times in passenger robust timetabling. In *Proceedings of the 7th International Seminar on Railway Operations Modelling and Analysis (RailLille2017), Lille, France. 8th best paper of the RailLille2017 conference*, 2017a.

Sofie Burggraeve, Simon H. Bull, Richard M. Lusby, and Pieter Vansteenwegen. Train turn restrictions and line plan performance. In *Proceedings of the Ninth Triennial Symposium on Transportation Analysis (TRISTAN IX), Oranjestad, Aruba*, 2016.

Sofie Burggraeve, Thijs Dewilde, Peter Sels, and Pieter Vansteenwegen. Improving passenger robustness by taking passenger numbers and recurring delays explicitly into account on the tactical level. In *Proceedings of the 6th International Seminar on Railway Operations Modelling and Analysis (RailTokyo2015), Tokyo, Japan*, 2015b.

Sofie Burggraeve, Sofie Van Thielen, and Pieter Vansteenwegen. Advanced methods for improving passenger robustness in railway systems. In *Proceedings of the Conference on Advanced Systems in Public Transport (CASPT2015), Rotterdam, The Netherlands*, 2015c.

Sofie Van Thielen, Sofie Burggraeve and Pieter Vansteenwegen. Optimal train reschedulling after conflict detection. In *Proceedings of the Conference on Advanced Systems in Public Transport (CASPT2015), Rotterdam, The Netherlands*, 2015.

Meeting abstracts, presented at international scientific conferences and symposia, published or not published in proceedings or journals

Sofie Burggraeve and Pieter Vansteenwegen. Optimizing buffer times and supplements in passenger robust timetabling. *31th Belgian Conference on Operations Research (ORBEL31), Brussels, Belgium*, February 2-3, 2017.

Sofie Burggraeve, Simon H. Bull, Richard M. Lusby and Pieter Vansteenwegen. Integration of railway line planning and timetabling. *30th Belgian Conference on Operations Research (ORBEL30), Louvain-la-Neuve, Belgium*, January 28-29, 2016.

Sofie Burggraeve, Simon H. Bull, Richard M. Lusby and Pieter Vansteenwegen. Integrating robust timetabling in railway line planning. *European Conference on Operational Research (EURO2016), Poznan, Polen*, July 3-6, 2016.

Sofie Burggraeve, Sofie Van Thielen and Pieter Vansteenwegen. Passenger Robust Train Planning in Complex Station Areas. *European Conference on Operational Research (EURO2015), Glasgow, United Kingdom*, July 12-15, 2015.

Sofie Van Thielen, Sofie Burggraeve and Pieter Vansteenwegen. Optimal train rescheduling after conflict detection. *29th Belgian Conference on Operations Research (ORBEL29), Antwerp, Belgium*, February 5-6, 2015.

Thijs Dewilde, Sofie Burggraeve, Tim Ameye, Peter Sels, Dirk Cattrysse and Pieter Vansteenwegen. The impact of preprocessing for solving the train routing problem. *28th Belgian Conference on Operations Researh (ORBEL28), Mons, Belgium*, January 30-31, 2014.

Awards

Sofie Burggraeve and Pieter Vansteenwegen. Optimization of supplements and buffer times in passenger robust timetabling. In *Proceedings of the 7th International Seminar on Railway Operations Modelling and Analysis (RailLille2017), Lille, France. 8th best paper of the RailLille2017 conference*, 2017a.

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF MECHANICAL ENGINEERING
KU LEUVEN MOBILITY RESEARCH CENTRE
Celestijnenlaan 300A box 2422
B-3001 Leuven
sofie.burggraeve@gmail.com
http://www.mech.kuleuven.be/en/cib